(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2002 (22.08.2002)

PCT

(10) International Publication Number
WO 02/064750 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/04909

(22) International Filing Date: 14 February 2002 (14.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/269,199	14 February 2001 (14.02.2001)	US
60/269,188	14 February 2001 (14.02.2001)	US
09/847,232	2 May 2001 (02.05.2001)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/847,232 (CIP)
Filed on 2 May 2001 (02.05.2001)

(71) Applicant *(for all designated States except US)*: **VENTRIA BIOSCIENCE** [US/US]; 4110 North Freeway Boulevard, Sacramento, CA 95834-1219 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **HUANG, Ning** [CN/US]; 1226 Caricia Drive, Davis, CA 95616 (US). **YANG, Daichang** [CN/US]; 1318 F Street, #3, Davis, CA 95616 (US).

(74) Agents: **ARNETT, Stephen, E.** et al.; Perkins Coie LLP, P.O. Box 2168, Menlo Park, CA 94026 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/064750 A2

(54) Title: EXPRESSION SYSTEM FOR SEED PROTEINS

(57) Abstract: The invention is directed to methods and compositions for high level expression of heterologous polypeptides in the grains of transgenic plants and compositions resulting therefrom that are suitable for oral delivery or other uses.

EXPRESSION SYSTEM FOR SEED PROTEINS

Field of the Invention

The present invention relates to the production of heterologous polypeptides in the grains of transgenic plants, methods, vectors and transformed hosts for producing the same, and compositions comprising such polypeptides and nucleic acids.

References

The following references are cited herein, and to the extent they may be pertinent to the practice of the invention, are incorporated herein by reference.

Alber and Kawasaki, *Mol. and Appl. Genet.* 1:419-434, 1982.

Aniansson, G. *et al.* (1990). "Anti-Adhesive Activity Of Human Casein Against Streptococcus Pneumoniae And Haemophilus Influenzae," *Microb Pathog*, 8(5):315-323.

Arnold R.R. *et al.*, *Infect Immun.* 28:893-898, 1980.

Ausubel F.M. *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, N.Y., 1993.

Bhan, M.K. *et al.*, *J Pediatr Gastroenterol Nutr* 7:208-213, 1988.

Beatty, K., *et al.*, "Kinetics of association of serine proteinases with native and oxidized alpha-1-proteinase inhibitor and alpha-1-antichymotrypsin," *J. Biol. Chem.* 255:3931-3934, 1980.

Boesman-Finkelstein M. and Finkelstein R.A. *FEBS Letters*, 144:1-5, 1982.

Bradford M. *Analytical Biochem.* 72:248-254, 1976.

Brandt, *et al.*, *Carlsberg Res. Commun.* 50:333-345, 1985.

Briggs, D, *et al.* (1981). Malting And Brewing Science - Volume 1, Malt And Sweet Wort (Chapman & Hall, New York, 2nd edition).

Briggs, D (1998). Malts And Malting (Blackie Academic And Professional, New York).

Bullen J.J. *et al.*, *Br. Med. J.* 1, 69-75, 1972.

Carrell, R.W. *et al.* "Structure and variation of human alpha-1-antitrypsin", *Nature* 298:329-334, 1983.

Castañón M.J. *et al.*,. *Gene*, 66:223-234, 1988.

Chandan R.C., *J Dairy Sci*, 51:606-607, 1968.

Chong, D.K. and Langridge, W.H., *Transgenic Res* 9:71-8, 2000.

Chowanadisai, W. and Lönnerdal, B., "Alpha-1-antitrypsin and antichymotrypsin in human milk: origin, concentrations, and stability", *Am. J. Clin. Nutr.* 2001 (in press).

Chrispeels K., *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53, 1991.

Clark *et al., J. Biol. Chem.* 264:17544-17550, 1989.

Davidson LA, *et al.* (1987). "Persistence Of Human Milk Proteins In The Breast-Fed Infant", *Act Pediatric Scand* 76(5):733-740.

Davidson, L.A., and Lönnerdal, B. "Fecal alpha-1-antitrypsin in breast-fed infants is derived from human milk and is not indicative of enteric protein loss", *Acta Paediatr. Scand.* 79:137-141, 1990.

della-Cioppa *et al., Plant Physiol.* 84:965-968, 1987.

Dellaporta S.L. *et al., Plant Mol. Biol. Rep.* 1:19-21, 1983.

Depicker *et al., Mol. Appl. Genet.* 1:561-573, 1982.

Dewey K.G. *et al., J Pediatrics* 126:696-702, 1995.

Dewey K.G. *et al., Pediatrics* 89:1035-1041, 1992.

Dewey K.G. *et al., Am J Clin Nutr* 57:140-145, 1993.

Ditta *et al., Proc. Nat. Acad. Sci., U.S.A.* 77:7347-7351, 1980.

Doncheck, J (1995). *Malts And Malting.* In: Kirk-Othmer Encyclopedia Of Chemical Technology (John Wiley & Sons, New York, 4th edition), volume 15, pp. 947-962.

Eley, E, (1990), *A Wandful World.* Food Processing August 1990, pp. 17-18

Faure A. and Jollès P., *et al., Comptes Rendus Hebdomadaires des Seances de L Academie des Sciences. D: Sciences Naturelles*, 271:1916-1918, 1970.

Fujihara T. and Hayashi K., *Archives of Virology* 140:1469-1472, 1995.

Gastañaduy, A. *et al., J Pediatr Gastroenterol Nutr* 11:240-6, 1990.

Gelvin, S.B. *et al.*, eds. PLANT MOLECULAR BIOLOGY MANUAL, 1990.

Gielen, *et al., EMBO J.* 3:835-846, 1984.

Goto, F *et al., Nature Biotech.* 17:282-286, 1999.

Grover M. *et al., Acta Paediatrica* 86:315-316, 1997.

Hamosh, M, *et al.* (1999). "Protective Function Of Human Milk: The Fat Globule," *Semin Perinatol* 23(3):242-9.

Harmsen M.C. *et al., Journal of Infectious Diseases* 172:380-388, 1995.

Hickenbotttom, JW (1977). *Sweeteners In Biscuits And Crackers.* The Bakers Digest, December 1977, pp. 18-22.

Hickenbottom, JW (1983). AIB Research Department Technical Bulletin, V(3), March 1983, pp. 1-8.

Hickenbottom, JW (1996). *Processing, Types, And Uses Of Barley Malt Extracts And Syrups.* Cereal Foods World 41(10): 778-790.

Hickenbottom JW (1997). *Malts In Baking.* American Society Of Bakery Engineers, Technical Bulletin #238, August 1997, pp. 1010-1013.

Huang N. *et al.*, *Plant Mol. Biol.*, 23:737-747, 1993.

Hough, JS, *et al.* (1982).). Malting And Brewing Science - Volume 2, Hopped Wort And Beer (Chapman & Hall, New York, 2nd edition).

Huang N. *et al.*, *J CAASS*, 1:73-86, 1990a.

Horvath H. *et al.*, *Proc. Natl. Acad. Sci. USA*, 97:1914-1919, 2000.

Humphreys, D.P. and Glover, D.J. "Therapeutic antibody production technologies: molecules, applications, expression and purification," *Curr Opin Drug Discov Devel* Mar;4(2):172-85, 2001.

Huang N. *et al.*, *Plant Mol. Biol.* 14:655-668, 1990b.

Huang, J., *et al.*, "Expression and purification of functional human alpha-1-antitrypsin from cultured plant cells", *Biotechnol. Prog.* 17:126-133, 2001.

Jensen L.G. *et al.*, *Proc. Natl. Acad. Sci. USA*, 93:3487-3491, 1996.

Jigami Y. *et al.*, *Gene*, 43:273-279, 1986.

Johnson, T. "Human alpha-1-proteinase inhibitor", *Methods Enzymol.* 80:756-757, 1981.

Jollès P., LYSOZYMES--MODEL ENZYMES IN BIOCHEMISTRY AND BIOLOGY. Birkhäuser Verlag, Basel; Boston, 1996.

Jouanin *et al., Mol. Gen. Genet.* 201:370-374, 1985.

Kortt A.A. *et al.*, "Dimeric and trimeric antibodies: high avidity scFvs for cancer targeting," *Biomol Eng* 18:95-108, 2001.

Kovar M.G. *et al.*, *Pediatrics* 74:615-638, 1984.

Kunz C. *et al.*, *Clin. Perinatol.* 26(2):307-33, 1999.

Lake, C, (1988). *From Strength To Strength.* Flood Flavorings, Ingredients, Packaging And Processing, February 1988, pp. 55 -57.

Langridge *et al.*, *Planta* 156:166-170, 1982.

Larrick, J.W. *et al.*, "Production of antibodies in transgenic plants", *Res Immunol* 149:603-8, 1998.

Lásztity, R. The chemistry of cereal proteins. CRC Press, Boca Raton, 1996.

Lee-Huang S. *et al.*, *Proc. Natl. Acad. Sci. USA,* 96:2678-2681, 1999.

Lindberg, T. "Protease inhibitors in human milk", *Pediatr. Res.* 13:969-972, 1979.

Lollike K. *et al.*, *Leukemia*, 9:206-209, 1995a.

Lönnerdal B., *Am L Clin Nutr*, 42:1299-1317, 1985.

Lönnerdal, B. "Recombinant milk proteins- an opportunity and a challenge", *Am. J. Clin. Nutr.* 63:622S-626S, 1996.

Maga E. *et al.*, *J. of Food Protection*, 61:52-56, 1998.

Maga E. *et al.*, *Transgenic Research*, 3:36-42, 1994.

Maga E. *et al.*, *Journal of Dairy Science*, 78:2645-2652, 1995.

Maniatis, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Edition, 1989.

Matsumoto, A. *et al.*, *Plant Mol Bio* 27:1163-72, 1995.

Maynard, J. and Georgiou, G. Antibody engineering, Annu Rev Biomed Eng 2:339-76, 2000.

McBride and Summerfelt, *Plant Mol. Biol.* 14:269-276, 1990.

Mcgilligan, KM, *et al.*, (1987), "Alpha-1-Antitrypsin Concentration In Human Milk," *Pediatr Res* 22(3):268-27o.

Mitra A. and Zhang Z., *Plant Physiol.* 106:977-981, 1994.

McGilligan, K.M. *et al.*, "Alpha-1-antitrypsin concentration in human milk," *Pediatr. Res.* 22:268-270, 1987.

Mitra, A, and Zhang, Z., *Plant Physiol* 106:977-81, 1994.

Moe, T, *et al.* (1995). *Mashing Of Rice With Barley Malt Under Non-Conventional Process Conditions For Use In Food Processes.* International Journal Of food Science And Technology 29:635-649.

Moore, K (1978). "Rice-Barley Malt Combo Improves Functionality And Nutrition Of Meats, Breads, And Baked Goods", *Food Product Development* 12(5):74.

Moreira-Ludewig, R. and Healy, C., *J. Pharm. and Tox. Meth.*, 27:75-100, 1992.

Motil K.J., *Curr Opin Pediatr* 12(5):469-76, 2000.

Murphy MS (1998). "Growth Factors And The Gastrointestinal Tract", *Nutrition* 14(10):771-4.

Nakajima H. *et al.*, *Plant Cell Reports*, 16:674-679, 1997.

Newburg, DS (1999). "Human Milk Glycoconjugates That Inhibit Pathogens", *Curr Med Chem* 6(2):117-127.

Newburg, DS, *et al.* (1998). "Role Of Human-Milk Lactadherin In Protection Against Symptomatic Rotavirus Infection", *Lancet* 351(1910):1160-4.

NIH publication, American Academy of Allergy and Immunology Committee on adverse reaction to food and National Institute of Allergy and Infectious Diseases, NIH, Bethesda, 1984.

Peeters, K., *et al.*, "Production of antibodies and antibody fragments in plants", *Vaccine* 9:2756-61, 2001.

Peterson JA, *et al.* (1998). "Glycoproteins Of The Human Milk Fat Globule In The Protection Of The Breast-Fed Infant Against Infections", *Bio Neonate* 74(2):143-162.

Piper, D.W. and Fenton, B.H. "pH stability and activity curves of pepsin with special reference to their clinical importance", *Gut* 6:506-508, 1965.

Prosser, CG (1996). "Insulin-Like Growth Factors In Milk And Mammary Gland," *J Mammary Gland Bio Neoplasia* (3):297-306.

Raikhel N., *Plant Phys.* 100:1627-1632, 1992.

Rey, M.W., *Nucleic Acid Res.* 18"5288, 1990.

Romer *et al., Biochem. Biophys. Res Commun.* 196:1414-1421, 1993.

Rudloff, S. and Lönnerdal, B. "Solubility and digestibility of milk proteins in infant formulas exposed to different heat treatments", *J. Pediatr. Gastroenterol. Nutr.* 15:25-33, 1992.

Saarinen K.M. *et al.*, *Adv Exp Med Biol* 478:121-30, 2000.

Salmon V. *et al.*, *Protein Expression and purification* 9:203-210, 1997.

Salmon, V. *et al.*, Protein Expr Purif 13:127-135, 1998.

Samaranayake Y.H. *et al.*, *Apmis*, 105:875-883, 1997.

Sambrook J. *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL (Second Edition), Cold Spring arbor Press, Plainview, N.Y, 1989.

Satue-Gracia M.T. *et al.*, *J Agric Food Chem* 48(10):4984-90, 2000.

Schütte H. and Kula M.R., *Biotechnology and Applied Biochemistry*, 12:599-620, 1990.

Sfat, M, et al (1981). *Malts And Malting.* In: Kirk-Othmer Encyclopedia Of Chemical Technology (John Wiley & Sons, New York, 3rd edition), volume 14, pp. 810-823.

Shah *et al., Science* 233:478-481, 1986.

Shin , K, *et al.* (2000). "PCR Cloning And Baculovirus Expression Of Human Lactoperoxidase And Myeloperoxidase", *Biochem Biophys Res Commun* 27(13):831-836.

Shugar D., *Biochim. Biophys. Acta*, 8:680-68, 1952.

Symbicon AB (1995). "Human Milk Kappa-Casein And Inhibition Of Helicobacter pylori Adhesion To Human Gastric Mucosa", *J Pediatr Gastroenterol Nujtr* 21(3):288-296.

Takai I. *et al.*, *J. Chrom. B, Biomedical Applications*, 685:21-25, 1996.

Tsuchiya K. *et al.*, *Applied Microbiology and Biotechnology*, 38:109-114, 1992.
Von Heijne *et al.*, *Plant Mol. Biol. Rep.* 9:104-126, 1991.
Wang C.S. and Kloer H.U., *Anal. Biochem.*, 139:224-227, 1984.
Wang C. *et al.*, *Comp. Biochem. Physiol.* 78B:575-580, 1984.
Ward P.P. *et al.*, *Bio/technology* 10:784-789, 1992.
Ye X. *et al.*, *Science* 287:303-305, 2000.
Yoshimura K. *et al.*, *Biochem. Biophys Res Com*, 150:794-801, 1988.

Background Of The Invention

Recombinant proteins have been expressed *in vitro* in different host expression systems such as bacterial cells, yeast and other fungi, mammalian cells, insect cells and, to a certain extent, plants. Each host expression system has its associated advantages and disadvantages.

Plants are attractive as hosts for expression of recombinant proteins as they are free from animal viruses and from toxins that are sometimes associated with microbial hosts. Scale-up can be performed much more easily simply by planting more acreages. Further, to the extent that the plant system is edible, recombinant molecules expressed in plant hosts may not require substantial purification if the recombinant molecules can retain bioactivity upon being ingested. However, up to the present, the level of expression heterologous proteins in transgenic plants has been low and purification of recombinant proteins from portions of the plant, such as leaves, etc., can be costly, making such an expression system commercially impractical.

There is, thus, a need for a reliable method and system for effecting high level expression of recombinant or heterologous polypeptides in plants. Such a system may require a unique or novel combination of components parts such as one or more of: promoter, enhancer, transcription factor, codon-optimized heterologous gene, terminator, leader sequences, selectable marker, etc., that can operate efficiently together.

Wu, C.Y. *et al.*, *Plant Cell Physiol.* 39(8):885-889, 1998 compares the activity of promoters from rice seed storage proteins using the 5'-flanking regions of two glutelin genes, *GluA-3* and *GluB-1*; a rice prolamin gene, *NPR33*; a rice α-globulin gene, *Glb*; and a rice allergenic protein gene, *RAG-1*; fused transcriptionally or translationally to β-glucuronidase ("GUS") reporter gene followed by a Nos terminator. Wu *et al.* concluded that, in their hands, "[o]n the basis of averaged values, the 13 kDa prolamin *NPR33* and the α-globulin *Glb* promoters directed the highest GUS gene expression, whereas the

allergenic protein albumin RAG-1 exhibited the lowest promoter activity. The glutelin GluA-3 promoter displayed an intermediate level activity." (Page 886)

International application WO 99/16890 discloses experiments relating to, among other things, the transformation of barley with nucleic acid constructs carrying maize embryo-specific globulin *Glb1* promoter, maize ubiquitin *Ubi1* promoter, barley endosperm-specific B_1 hordein promoter or the barley endosperm-specific D-hordein promoter, with the reporter *uidA* gene (encoding GUS activity). The application also discloses the use of a B_1 hordein signal sequence in conjunction with the B_1 hordein promoter for driving the expression of the *uidA* gene and concluded that, in their hands, the use of the B_1 hordein signal sequence enhanced expression of the *uidA* gene under the control of the B_1 hordein promoter in barley.

There is a continue need in the art for expression systems including methods, vectors, hosts and other compositions, that are capable of effecting higher level expression of functional recombinant polypeptides in plants.

Summary Of The Invention

It is one of the objects of the present invention to address the unmet need for reliable methods for high level expression of heterologous polypeptides in plants.

It is another one of the objects of the present invention to provide vectors, hosts, and methods for such expression and compositions containing such.

In one aspect, the invention includes method of producing a heterologous polypeptide in a grain (seed) of a plant. The method includes culturing a transformed plant to form a grain-producing transformed plant; and recovering transgenic grains containing the heterologous polypeptide from the grain-producing transformed plant. The transformed plant contains a chimeric gene having (i) a rice glutelin Gt1 promoter or a rice globulin Glb promoter, (ii) a nucleic acid leader encoding a rice glutelin Gt1 signal peptide; and (iii) a heterologous nucleic acid that encodes the heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide.

The plants are obtained by transforming plant cells; culturing the transformed plant cells to form a plurality of grain-producing transformed plants; and selecting a transformed plant that exhibits high level expression of the heterologous polypeptide in its grains.

The transformed plant may further include a transcription factor-encoding nucleic acid that encodes one or more heterologous transcription factor that is capable of

enhancing expression of the heterologous polypeptide. Exemplary transcription factors include O2 or PBF, Reb, or active fragment thereof.

The plant may be a dicot or monocot, and is preferably a monocot such as rice, wheat or barley. The heterologous protein produced in monocots is preferably a non-plant storage polypeptide, such as an antibody, cytokine, lymphokine, chemokine, polypeptide hormones growth factor, coagulation factor, anti-infective, or cytotoxin, a non-human animal polypeptide, a milk protein, an anti-inflammatory protein, an intestinal trefoil factor (ITF) or an active fragment thereof, and more particularly, lactoferrin, lysozyme, lactoferricin, intestinal trefoil factor (ITF), epidermal growth factor (EGF), keratinocyte growth factor (KGF), insulin-like growth factor I (IGF-I), lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin, immunoglobulins, alpha-lactalbumin, beta-lactoglobulin, alpha-casein, beta-casein, albumin, fibrinogen, or a protease inhibitor.

Exemplary chimeric genes include (a)a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized lysozyme coding sequence; (b) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized lactoferrin coding sequence; (c)a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized EGF; (d) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized IGF-I; (e) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized haptocorrin; (f) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and codon optimized IGF-I; (g) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized lactahedrin;(h)a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized lactoperoxidase; and (i) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized α_1-antitrypsin.

An exemplary chimeric gene is composed of (a) a rice globulin Glb promoter, a rice glutelin Gt1 signal peptide, and codon optimized haptocorrin; or (b) a rice globulin Glb promoter, a rice glutelin Gt1 signal peptide, and codon optimized lysozyme.

The transformed plant may include a second chimeric gene, which includes (i) an aleurone-specific promoter; and (ii) the heterologous nucleic acid encoding the heterologous polypeptide, operably linked together to enable expression of the heterologous polypeptide.

In another aspect, the invention includes an expression vector that contains the above chimeric gene, and preferably, also a selectable marker gene.

In still another aspect, the invention includes a transformed host containing this expression vector. the host being a plant cell, a plant or a grain. The transformed host may also be transformed with one or more coding sequences encoding one or more heterologous transcription factors capable of enhancing expression of the heterologous polypeptide in the transformed host.. The plant may be a monocot or a dicot, preferably a monocot such as rice, barley, and wheat.

In another aspect, the invention includes transformed grains containing the chimeric gene described above.

In still another aspect, the invention includes a codon optimized nucleic acid molecule for expression of the polypeptides: (i)a h LF nucleic acid encoding human lactoferrin; as exemplified by SEQ ID NO: 3; (ii) a hLZ nucleic acid encoding human lysozyme; as exemplified by SEQ ID NO: 1, (iii) a hEGF nucleic acid encoding human epidermal growth factor; as exemplified by SEQ ID NO: 8; and (iv) hIGF-I nucleic acid encoding human insulin-like growth factor , as exemplified by SEQ ID NO: 9.

Also disclosed is a method for producing a heterologous polypeptide in a grain of a plant, comprising the steps of culturing a transformed plant to form a grain-producing transformed plant; and recovering grains containing the heterologous polypeptide from the grain-producing transformed plant. The transformed plant has a first chimeric gene and at least one heterologous transcription factor that is capable of enhancing the expression of the first chimeric gene. The chimeric gene comprises: (i) a promoter of a storage protein gene; (ii) a nucleic acid leader encoding a signal peptide of a storage protein; and (iii) a heterologous nucleic acid that encodes the heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide. The transformed plant may further comprises a second chimeric gene having (i) an aleurone specific promoter; and (ii) the heterologous nucleic acid, all operably linked to enable expression of the heterologous polypeptide.

An exemplary storage protein gene may be selected from the group consisting of: rice glutelins, rice oryzins, rice prolamines, barley hordeins, wheat gliadins, wheat glutenins, maize zeins, maize glutelins, oat glutelins, sorghum kafirins, millet pennisetins, and rye secalins. Preferred promoters are a rice Gt1 promoter, a rice Glb promoter or a wheat glutenin promoter. An exemplary signal peptide is also derived from the same gene, preferably a rice Gt1 signal peptide or rice Glb signal peptide. An

exemplary heterologous transcription factor may be selected from the group consisting of Reb, O2, and PBF.

An exemplary aleurone specific promoter is the promoter of the lipid transfer protein gene Ltp1 or a promoter of chitinase gene Chi26.

The transformed plant may be a dicot or monocot, preferably a cereal monocot such as rice, wheat, or barley.

As above, the heterologous protein produced in monocots is preferably a non-plant storage polypeptide, such as an antibody, cytokine, lymphokine, chemokine, polypeptide hormones growth factor, coagulation factor, anti-infective, or cytotoxin, a non-human animal polypeptide, a milk protein, an anti-inflammatory protein, an intestinal trefoil factor (ITF) or an active fragment thereof, and more particularly, lactoferrin, lysozyme, lactoferricin, intestinal trefoil factor (ITF), epidermal growth factor (EGF), keratinocyte growth factor (KGF), insulin-like growth factor I (IGF-I), lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin, immunoglobulins, alpha-lactalbumin, beta-lactoglobulin, alpha-casein, beta-casein, albumin, fibrinogen, or a protease inhibitor.

In another aspect, the invention includes a first chimeric gene that comprises (i) an aleurone-specific promoter and (ii) a heterologous nucleic acid encoding a heterologous polypeptide, operably linked to enable expression of the heterologous polypeptide. The vector may further include a transcription factor-encoding nucleic acid that encodes at least one heterologous transcription factor that is capable of enhancing expression of the heterologous polypeptide. The vector may further include (i) a promoter of a storage protein gene; (ii) a nucleic acid leader encoding a signal peptide of a storage protein; and (iii) the heterologous nucleic acid encoding the heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide.

Also disclosed is a transformed host comprising a first chimeric gene comprising an aleurone-specific promoter operably linked to a first heterologous polypeptide to enable expression of the heterologous polypeptide. The host may further have at least one heterologous transcription factor, and a second chimeric gene comprising (i) a promoter of a storage protein gene, (ii) a nucleic acid leader encoding a signal peptide of a storage protein, and (iii) a heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide. Also disclosed is a heterologous polypeptide in the above seed .

A method of processing a transgenic seed that comprises a heterologous polypeptide, in accordance with another aspect of the invention, includes the steps of

(a) powderizing the transgenic seed in a buffer to form bufferized powdered seed; (b)removing cell debris from the bufferized powdered seed to obtain a supernatant; and (c)reducing liquid from the supernatant to produce an extract that comprises the heterologous polypeptide. The buffer may be a volatile buffer such as ammonium bicarbonate buffer, ammonium acetate buffer.

Another method of processing seeds, in accordance with the invention, includes the steps of (a) providing transgenic grains that contain starch and heterologous polypeptides; (b)providing an enzymatic composition for conversion of malt to wort; (c) combining the transgenic grains with the enzymatic composition under conditions that allow for conversion of starch in the transgenic grains to sugars; (d)allowing the starch in the transgenic grains to be at least partially converted; and (e)separating resulting transgenic malt syrup from resulting transgenic grain residue, each containing the heterologous polypeptide. The enzymatic composition may include amylases, e.g., from a liquid malt extract comprising barley.

In a related aspect, the invention includes a method of processing transgenic grains containing a heterologous polypeptide to produce a malt extract. The method includes the steps of (a) providing transgenic grains that contain starch and heterologous polypeptides; (b) allowing the transgenic grains to undergo a malting process to form malted transgenic grains; and (c)crushing the malted transgenic grains.

The method may further include mixing the crushed and malted transgenic grains with water to obtain a liquid transgenic malt extract. Preferred seeds are cereal-grain seeds, such as corn, sorghum, rice, barley, rye, wheat, oats and triticale.
In accordance with another aspect of the invention, a transgenic malt syrup or extract comprising a heterologous polypeptide produced in transgenic grains in a malt syrup.

Also disclosed is an edible food comprising a heterologous polypeptide in a transgenic malt syrup, a transgenic malt extract, or a transgenic grain residue in a food product.

A protein-containing product of the invention includes a monocot seed flour, extract or malt composition containing and one or more seed-produced, non-seed proteins in substantially unpurified form, and a vehicle containing the composition in a form suitable for human or animal use. The vehicle is selected from the group consisting of a capsule, binder components effective to tabletize the composition, a consumable liquid, and a consumable suspension. The vehicle may be a processed food in which the composition is mixed.

Exemplary seed-produced, non-seed proteins include:

(i) lysozyme and lactoferrin, where the vehicle allows for ingestion;

(ii) a blood-clotting factor, where the vehicle is a surgical dressing, powder or cream.

(iii) an industrial enzyme, where the extract is in a liquid or powder form, and the vehicle is a liquid or powder dispenser, respectively.

(iv) a detergent enzyme, where the vehicle is a detergent,

(v) an anti-infective, and the vehicle is selected from the group consisting of: toothpaste, mouthwash, hand soap, dish soap, dishwasher detergent, tile cleaner, bathroom cleaner, toilet cleaner.

These and other objects and features of the invention will become more fully apparent when the following detailed description of the invention is read in conjunction with the accompanying figures.

Brief Description of the Figures

FIG. 1 is a map of the pAPI159 expression construct that contains the human lysozyme coding sequence under the control of a Gt1 promoter and Gt1 signal sequence.

FIG. 2 shows the results of Western blot analysis for the expression of recombinant human lysozyme in various tissues of rice plants, where lanes 1 and 15 are a human milk lysozyme standard; lane 2 is a broad range molecular weight marker from Sigma; lanes 3 and 4 represent mature seed tissue extracts; lanes 5 and 6 represent germinated seed extracts; lanes 7 and 8 represent root tissue extracts; lanes 9 and 10 represent extracts from young root tissue; lanes 11 and 12 represent leaf extracts; and lanes 13 and 14 represent extracts from young leaf; from untransformed ("U") or transgenic ("T") plants, respectively. The total loading protein amount was 40 µg per lane.

FIG. 3 shows the effect of incubating recombinant human lysozyme from transgenic rice seed, a human lysozyme standard (30 µg/ml), a control (20 mM sodium phosphate, pH 7.0, 5 mM EDTA) or an untransformed rice extract on the growth of *E. coli* strain JM109. At the end of the incubation (for the time indicated), an aliquot of the mixture was plated on LB plates and colony forming units per ml (CFU/ml) was calculated.

FIG. 4 is a graph showing the specific activity of lysozyme, as determined by incubating an identical concentration of a human lysozyme standard, human lysozyme from transgenic rice (plant) and lysozyme from chicken egg white with a standard

amount of *M. luteus*, followed by evaluation of the reduction in the turbidity due to the activity of lysozyme over five minutes.

FIG. 5A show thermal stability of human lysozyme ("Hlys") and recombinant human lysozyme from transgenic rice ("rHLys"). Lysozyme was dissolved at 100 µg/ml in PBS. The mixtures were subjected to different temperatures for different lengths of time. At the end of each heat treatment, the remaining lysozyme activity was assessed by activity assay.

FIG. 5B shows pH stability of Hlys and rHlys. Lysozyme was dissolved in different buffers at 100 µg/ml. The mixture was incubated at 37°C for 30 min. The lysozyme activity was determined by activity assay.

FIG. 6 presents the results of an analysis of lysozyme expression in transgenic rice grains over several generations. Proteins from 1 g of brown rice flour were extracted with 40 ml of extraction buffer containing 0.35 M NaCl in PBS. Extraction was conducted at room temperature for 1 h with shaking. Homogenate was centrifuged at 14,000 rpm for 15 min at 4°C. Protein supernatant was removed and diluted as needed for lysozyme turbidimetric activity assay. Extraction was repeated three times and standard deviation was shown as an error bar. Lysozyme yield was expressed as percentage of total soluble protein (%TSP).

FIG. 7 is a restriction map of the pAPI164 plasmid that contains the human lactoferrin coding sequence under the control of a rice glutelin (Gt1) promoter, aGt1 signal peptide, and a nopaline synthase (NOS) terminator/polyadenylation site.

FIG. 8 shows the results of a SDS-PAGE analysis for human lactoferrin stained with Coomassie blue, where lane 1 is the molecular weight marker; lanes 2 - 5 are purified human derived lactoferrin (Sigma, USA); lanes 6 - 10 are single seed extracts from homozygous transgenic lines and lane 11 is a seed extract from non-transformed TP-309.

FIG. 9 shows the results of a Western blot analysis of various tissues of the transgenic rice plants, demonstrating the tissue specificity of rLF expression. Lane 1 is the molecular weight marker; lane 2 is human lactoferrin (Sigma, USA); lane 3 is an extract from leaf; lane 4 is an extract from sheath; lane 5 is an extract from root; lane 6 is an extract from seed and lane 7 is an extract from 5-day germinated seeds.

FIG. 10 is a bar diagram illustrating the bactericidal effect of native human lactoferrin ("nHLF") and purified recombinant human lactoferrin produced by transgenic rice ("rHLF") on growth of *E. coli* (EPEC) after pepsin/pancreatic treatment.

FIG. 11 is a graph illustrating pH-dependent iron release by native human lactoferrin ("nHLF") and purified recombinant human lactoferrin produced by transgenic rice seeds ("rHLF").

FIG. 12 shows the binding and uptake of HLf to Caco-2 cells after *in vitro* digestion. FIG. 12 A shows the determination of Dissociation constant. FIG. 12B shows the number of binding sites for HLf on Caco-2 cells. FIG. 12C shows the total uptake of HLf and Fe to Caco-2 cells within 24 h. FIG. 12D shows degradation of HLf after uptake into Caco cells determined by the amount of free ^{125}I in the cell fractions.

FIG. 13 shows three AAT plasmids: pAPI255 containing Glb promoter, Glb signal peptide, codon-optimized AAT gene, Nos terminator and ampicillin resistance gene; pAPI250 containing Gt1 promoter, Gt1 signal peptide, codon-optimized AAT gene, Nos terminator and ampicillin resistance gene; and pAPI282 containing Bx7 promoter, Bx7 signal peptide, codon-optimized AAT gene, Nos terminator and ampicillin resistance gene.

FIG. 14 shows Coomassie brilliant blue staining of aqueous phase extraction of transgenic rice cells expressing human AAT. Both untransformed and transgenic rice grains were ground with PBS. The resulting extract was spun at 14,000 rpm at 4° C for 10 min. Supernatant was collected and loaded onto a precast SDS-PAGE gel.

FIG. 15 shows Western blot analysis of recombinant human AAT from transgenic rice grains. The extract from transgenic rice grain was separated by SDS-PAGE gel and then blotted onto a filter. The identification of AAT in rice grain was carried out by anti-AAT antibody by Western analysis. FIG. 16 shows Coomassie staining (FIG. 16A) and western blot analysis (FIG. 16B) of protein from transgenic rice grains expressing AAT. The activity of rAAT was demonstrated by a band shift assay. AAT samples from different sources were incubated with equal moles of porcine pancreatic elastase (PPE) at 37^{0}C for 15 min. Negative control for band shift assay was prepared with the AAT samples incubated with equal volume of PPE added. Lane M is molecular weight markers. Lane 1a is purified AAT from human plasma. Lane 1b is purified AAT from human plasma + PPE. Lane 2a is protein extract containing AAT from transgenic rice seed; Lane 2b is protein extract containing AAT from transgenic rice seed + PPE. Lane 3a is untransformed seed extract. Lane 3b is untransformed seed extract + PPE. A shifted band was shown in lane 1b, 2b and 3b in FIG. 16A. The shifted band was confirmed to contain AAT entity by Western blot in FIG. 16B.

FIG. 17A-C are schematic representations of 3 plasmids containing the Reb coding sequence under the control of 3 different promoters. FIG. 17A shows the

globulin promoter (Glb), with the Reb gene and the Reb terminator. FIG. 17B shows the actin promoter (Act), with the Reb gene and the Reb terminator. FIG. 17C shows the native Reb promoter, with the Reb gene and the Reb terminator.

FIG. 18A-B are schematic depictions of 2 plasmids which contain different transcription factor coding sequences under the control of the rice endosperm-specific glutelin promoter (Gt-1). FIG. 18A shows plasmid pGT1-BPBF (API286) containing the Gt1 promoter, barley prolamin box binding factor (BPBF), Nos terminator and kanamycin resistance gene. FIG. 18B shows pGT1-PBF (API285) containing the Gt1 promoter, the maize prolamin box binding factor (PBF), Nos terminator and kanamycin resistance gene.

FIG. 19 illustrates the results of an analysis for the expression of recombinant human lysozyme in mature seed of T_0 transgenic plants derived from progenitor cells transformed with constructs containing the human lysozyme gene expressed under the control of the Glb promoter and the Reb gene expressed under the control of its own promoter ("Native-Reb"). Seeds of 30 plants containing the Reb and lysozyme genes and seeds from 17 plants containing only the lysozyme gene were analyzed for lysozyme, with twenty individual seeds of each plant analyzed.

FIG. 20 is a comparison of the codon-optimized epidermal growth factor sequence ("Egfactor") with a native epidermal growth factor sequence ("Native Gene"), aligned to show 53 codons in the mature sequences, with 27 (51%) codon changes and 30 (19%) nucleotides changes.

FIG. 21 is a restriction map of the 4,143 bp plasmid, API270 (pGlb-EFG v2.1), showing an expression cassette for epidermal growth factor ("EGF"), and containing a Glb promoter, a Glb signal peptide, codon optimized EGF, a Nos terminator and an ampicillin resistance selectable marker.

FIG. 22 is a restriction map of the 3877 bp plasmid, API303 (pGt1-EGF v2.1), showing an expression cassette for epidermal growth factor (EGF), and containing a rice Gt1 promoter, a Gt1 signal peptide, codon optimized EGF, a Nos terminator and an ampicillin resistance selectable marker.

FIG. 23 is a Western blot analysis of recombinant human EFG ("rhEGF") in transgenic rice seed. Lane 1 shows a broad range of molecular weight markers. Lane 2 shows rhEGF expressed in yeast, loaded at 125 ng. Lanes 2 to 6 show rhEGF expressed from different transgenic rice seeds. Lane 7 is from seeds of control untransformed TP 309.

FIG. 24 is a comparison of the codon-optimized insulin-like growth factor I

sequence ("Insgfact") with a native human insulin-like growth factor I sequence ("native gene"), aligned to show 70 codons in the mature sequences, with 40 (57%) codon changes and 47 (22%) nucleotides changes.

FIG. 25 is a restriction map of the 3928 bp plasmid, API304 (pGt1-IFG v2.1), showing an expression cassette for insulin-like growth factor I ("IGF"), and containing a rice Gt1 promoter, a Gt1 signal peptide, codon optimized IGF, a Nos terminator and an ampicillin resistance selectable marker.

FIG. 26 is a restriction map of the 4194 bp plasmid, API271 (pGlb-IGF v2.1), showing an expression cassette for insulin-like growth factor I ("IGF"), and containing a Glb promoter, a Glb signal peptide, codon optimized IGF, a Nos terminator and an ampicillin resistance selectable marker.

FIG. 27 is a Western blot analysis of recombinant human IGF-I ("rhIGF") expressed in transgenic rice seeds. Lane 1 shows a broad range of molecular weight markers. Lane 2 shows rhIGF expressed in yeast, loaded at 1 μg. Lanes 3-9 show rhIGF from different transgenic seeds. Lane 10 is from seeds of control untransformed TP 309.

FIG. 28 is a restriction map of the 5250 bp plasmid, API321 (pGlb-gt1sig-Haptocorrin v 2.1), showing an expression cassette for haptocorrin, and containing a Glb promoter, a Gt1 signal peptide, codon optimized haptocorrin, a Nos terminator and an ampicillin resistance selectable marker.

FIG. 29 is a restriction map of the 4948 bp plasmid, API320 (pGt1-Haptocorrin v 2.1), showing an expression cassette for haptocorrin, and containing a Gt1 promoter, a Gt1 signal peptide, codon optimized haptocorrin, a Nos terminator and an ampicillin resistance selectable marker.

FIG. 30 is a restriction map of the 4468 bp plasmid, API292 (pGlb-kcasein v2.1), showing an expression cassette for kappa-casein ("k-casein"), and containing a Glb promoter, a Glb signal peptide, a k-casein gene, a Nos terminator and an ampicillin resistance selectable marker.

FIG. 31 is a restriction map of the 4204 bp plasmid, API297 (pGT1-kaapa-Casein v2.1), showing an expression cassette for kappa-casein, and containing a Gt1 promoter, a Gt1 signal peptide, mature kappa-casein polypeptide encoding gene, a Nos terminator and an ampicillin resistance selectable marker.

FIG. 32 is a restriction map of the 4834 bp plasmid, API420 (pGt1-LAD), showing an expression cassette for lactahedrin, and containing a Gt1 promoter, a Gt1 signal peptide, lactohedrin gene, a Nos terminator and a kanamycin resistance

selectable marker.

FIG. 33 is a restriction map of the 5638 bp plasmid, API418 (pGT1-LPO-S), showing an expression cassette for lactoperoxidase (minus the propeptide), and containing a Gt1 promoter, a Gt1 signal peptide, lactoperoxidase gene without the propeptide, a Nos terminator and a kanamycin resistance selectable marker.

FIG. 34 is a restriction map of the 5801 bp plasmid, API416 (pGt1-lactoperoxidase), showing an expression cassette for codon optimized human lactoperoxidase, and containing a rice Gt1 promoter, a Gt1 signal peptide, codon optimized lactoperoxidase, a Nos terminator and a kanamycin resistance selectable marker.

FIG. 35 is a restriction map of the 4408 bp plasmid, API230 (pBX7-Lysozyme v2.1.1), showing an expression cassette for codon optimized lysozyme, and containing a BX-7 promoter, a Gt1 signal peptide, codon optimized lysozyme gene, a Nos terminator and an ampicillin resistance selectable marker.

FIG. 36A and B represent schematic diagrams of the map of 2 plasmids, API254 (FIG. 36A) and API264 (FIG. 36B) containing heterologous protein coding sequences under the control of the rice endosperm-specific globulin promoter (Glb), the Glb signal peptide, and Nos terminator. API254 contains the lactoferrin coding sequence, and API264 contains the human lysozyme coding sequence.

FIG. 37 is a restriction map of the 4271 bp plasmid, API225, showing an expression cassette for codon optimized lysozyme, and containing a GT-3 promoter, a Gt1 signal peptide, codon optimized lysozyme, a Nos terminator and an ampicillin resistance selectable marker.

FIG. 38 is a restriction map of the 4106 bp plasmid, API229, showing an expression cassette for codon optimized lysozyme, and containing a RP-6 promoter, a Gt1 signal peptide, codon optimized lysosyme, a Nos terminator and an ampicillin resistance selectable marker.

FIG. 39 is a comparison of the expression of lysozyme under Gt1 or Glb promoter with Gt1 signal peptide or Glb signal peptide. FIG. 39A is a schematic representation of plasmid API159 that contains Gt1 promoter, Gt1 signal peptide, a lysozyme gene and Nos terminator; plasmid API 228 that contains Glb promoter, Gt1 signal peptide, a lysozyme gene and Nos terminator; and plasmid API264 that contains Glb promoter, Glb signal peptide, a lysozyme gene and Nos terminator. FIG. 39B shows the activities of lysozyme in lysozyme-positive seeds produced in transgenic rice plants transformed with API159, API228 and API264. The seeds from multiple lines of

each construct were analyzed by the lysozyme activity assay. Individual seeds from each plant were analyzed. Seeds lacking detectable amounts of lysozyme were excluded. The activities of 20-lysozyme-positive seeds per plant, including both hemizygous and homozygous seeds were averaged. The average activities were plotted on the chart.

FIG. 40 shows the expression time course of human lysozyme during endosperm development in transgenic line. Ten spikelets were harvested at 7, 14, 21, 28, 35, 42 and 49 days after pollination ("DAP") and analyzed by the lysozyme activity assay. The dark bars were from 159-1-53-16-1. The light bars were from 264-1-92-6-1.

FIG. 41 is a bar graph comparing the level of lysozyme expression in transgenic T1 rice seeds under 7 different promoters: *Gt1*, *Glb*, *Glub-2*, *Bx7*, *Gt3*, *Glub-1* and *Rp6*. All constructs contained a *Gt1* signal peptide.

FIG. 42A-I are a double-stranded depiction of the DNA sequence of the rice (Oryza sativa) bZIP protein, designated ("Reb"). The gene sequence of 6.227 kb consists of 5 introns and 6 exons flanked by 1.2kb of the 5' promoter and 1.2 kb of the 3' region.

FIG. 43 presents a restriction map of the rice *Reb* gene isolated from BAC clone 42B9.

FIG. 44 is a single-stranded depiction of a portion of the DNA sequence of the Glb promoter with putative Reb binding sites indicated.

FIG. 45A presents a schematic diagram of the plasmid constructs used for transient assays illustrating the transactivation function of the *Reb* gene towards the *Glb* promoter where (1) is a GUS reporter construct with GUS expressed under the control of the *Glb* promoter (Glb-GUS-Nos); (2) is a null promoter construct (Δpromoter-Reb-Term) where the cells were co-bombarded with the Glb-GUS reporter gene construct; (3) is an Reb expression construct with *Reb* expressed under the control of the *Glb* promoter (Glb-Reb-Term), where the cells were co-bombarded with the *Glb*-GUS-Nos reporter gene construct; (4) is an Reb expression construct with *Reb* expressed under the control of the Actin promoter (Act-Reb-Term) where the cells were co-bombarded with the Glb-GUS-Nos reporter gene construct; and (5) is an Reb expression construct with *Reb* expressed under the control of the native *Reb* promoter (Native-Reb-Term), where the cells were co-bombarded with the Glb-GUS-Nos reporter gene construct.

FIG. 45B illustrates the relative GUS activity as measured in transient expression assays using the effector/reporter combinations shown in Fig. 45A.

FIG. 46A presents a schematic diagram of the plasmid constructs used for gain-of-function analysis Reb transactivation by transient assay where (1) is a GUS reporter construct with GUS expressed under the control of the *Gt1* promoter (Gt1-GUS-Nos); (2) is Gt1-GUS-Nos where the cells were co-bombarded with an expression construct which has Reb expressed under the control of the native Reb promoter (native-Reb-Term); (3) is a GUS reporter construct with GUS expressed under the control of the *Gt1* promoter modified to contain the Reb response sequence, UAS (Gt1-UAS-GUS-Nos); and (4) is modified Gt1-UAS-GUS-Nos where the cells were co-bombarded with native-Reb-Term.

FIG. 46B illustrates the relative GUS activity as measured in transient expression assays using the reporter constructs shown in FIG. 46A.

FIG. 47 is a single-stranded depiction of the DNA sequence of the rice (Oryza sativa) globulin promoter *Glb*, with putative binding sites for the O2 transcription factor and the prolamin box indicated in the figure.

FIG. 48 is a single-stranded depiction of the DNA sequence of the wheat *Bx7* promoter with putative binding sites for the O2 transcription factor and the prolamin box indicated in the figure.

FIG. 49 illustrates the results of transactivation of the *Glb, RP6, PG5a* or *Bx7* gene promoter by O2 and PBF in rice immature endosperms, where FIG. 49A is a schematic diagram of reporter and effector plasmids used in transient expression assays.

FIG. 49B presents the results of transient expression assays following particle bombardment of rice immature endosperms, where O2 and PBF were independently expressed under the control of the (ubiquitin) *Ubi* promoter (*Ubi:O2*) and (*Ubi:*PBF), respectively. In addition, the *Ubi:O2* and *Ubi:*PBF constructs were individually co-bombarded with Glb/GUS/NOS, RP6/GUS/NOS, PG5a/GUS/NOS or Bx7/GUS/NOS. All results are given relative to the GUS/LUX ratio of the Glb/GUS/NOS, *RP6*/GUS/NOS, PG5a/GUS/NOS or Bx7/GUS/NOS construct, respectively. Error bars represents the standard deviation of the mean value for at least five independent particle bombardments.

FIG. 50 illustrates the results of transactivation of the *Gt1* promoter by various amounts of O2 or PBF effector plasmid in rice immature endosperm. FIG. 50A is a schematic diagram of reporter and effector constructs used in transient expression assays. FIG. 50B presents the results of transient expression assays following particle bombardment of rice immature endosperms where O2 and PBF were independently

expressed under the control of the (ubiquitin) *Ubi* promoter (*Ubi:O2*) and (*Ubi*:PBF), respectively. In addition, the *Ubi:O2* and *Ubi*:PBF constructs were individually co-bombarded with Gt1/GUS/NOS. All results are given relative to the GUS/LUX ratio of Gt1/GUS/NOS construct. Error bar represents the standard deviation of the mean value for at least five independent particle bombardments.

FIG. 51 illustrates the results of transactivation of the rice glutelin (*Gt1)* gene promoter by the O2, PBF, o2-676 and PBFm transcription factors in rice immature endosperm. FIG. 50A is a schematic diagram of the reporter and effector constructs used in the transient expression assay. FIG. 50B presents the results of transient expression assays following particle bombardment of rice immature endosperms, with O2, PBF, o2-676 and PBFm expressed under the control of the 35S promoter (*35S:O2*), (*35S*:PBF); (*35S*: o2-676); and (*35S*:PBFm), respectively. In addition, the activation based on co-bombard of the combination of O2 plus PBF with Gt1/GUS/NOS was evaluated, as was the combination of o2-676 plus PBFm with Gt1/GUS/NOS. The pAHC18 (Ubi/LUC/NOS) plasmid was used as an internal control for all experiments. All results are given relative to the GUS/LUX ratio of Gt1/GUS/NOS construct. Error bar represents the standard deviation of the mean value for at least five independent particle bombardments.

FIG. 52 depicts the results of a PCR analysis of T_0 transgenic plants containing Reb and the human lysozyme gene. Fig. 52A is a diagram that shows the construct API266 (native-Reb), the primer positions and the size of the 522 bp amplified fragment, where one primer was designed based on the vector sequence and the other using the Reb terminator.

FIG. 52B (Reb) is a diagram that shows the construct API264 (Glb-lys), the primer positions and the size of the 278 bp amplified fragment. The primers hybridize to an internal sequence of the human lysozyme gene.

FIG. 52C presents the results of a PCR analysis of native-Reb/Glb-Lys co-transformed plants where arrows mark the 522 bp fragment of the Reb/vector region and the 278 bp fragment derived from the internal sequence of the human lysozyme gene.

FIG. 53A is a comparison of the PCR amplified 5′ flanking region of the *Chi26* gene (PCR-Chi26) with the sequences of *Chi26* (BLYChi26A*)* in GenBank (accession number: L34210). Asterisk (*) represents identical nucleotide. TATA box and E-region [Leah, R. et al., Plant J. 6:579-589 (1994)] are underlined. Transcriptional start site is in bold and underlined.

FIG. 53B is a schematic representation of the expression cassette (Chi26/GUS/NOS) under the control of the barley PCR-Chi26 promoter with a Nos terminator. The 882 bp PCR-Chi26 fragment was transcriptionally fused to the β-glucuronidase (GUS) gene.

FIG. 54A shows PCR fragment polymorphism of *Ltp1* promoter region among seven barley varieties. FIG. 54B is a comparison of the PCR-amplified 5′ flanking region of the *Ltp1* gene (PCR-Ltp1) with the sequence of *Ltp1* (BLYLTP1) in GenBank (accession number: X60292). TATA box and E-region [Leah, R. et al. (1994)] is underlined. Transcriptional start site is in bold and underlined. FIG. 54C is a schematic diagram of the expression cassette (Ltp1/GUS/NOS) under the control of the barley PCR-Ltp1 promoter. The 957 bp PCR-Ltp1 fragment was transcriptionally fused to the -glucuronidase (GUS) gene.

FIG. 55 is a Southern analyses of genomic DNA isolated from transgenic rice plants containing Ltp1/GUS/NOS or Chi26/GUS/NOS. Fig. 55A shows rice genomic DNA from Chi26/GUS/NOS transgenic lines: 217-1-1, 217-5-1, 217-8-1, and 217-12-1 (lanes 2-5), digested with *Hind*III and *Eco*RI, and probed with a 3 kb fragment containing the PCR-*Chi26* promoter, GUS coding region and the NOS terminator (FIG. 53B). Lane 1 is molecular weight (MW) marker (*/Hind*III). Lane 6 represents non-transformed TP309. Lane 7 and 8 show five (5X) and ten (10X) copies of a genome equivalent, respectively.

FIG. 55B Rice genomic DNA from transgenic lines: 220-1-1, 220-2-1, 220-11-, 220-13-1, and 220-14-1(lanes 2-6) was digested with *Hind*III and *Eco*RI and probed with a 2.9 kb fragment containing the PCR-Ltp1 promoter, GUS coding region and NOS terminator (FIG. 54C). Lane 1 indicates the molecular weight (MW) marker (*/Hind*III). Lane 7 indicates non-transformed TP309. Lane 8 and 9 show five (5X) and ten 10X) copies of a genome equivalent, respectively.

FIG. 56 is a restriction map of plasmid pAPI224 (pEM221-GUS v2.1) that contains the *Emp1* promoter, GUS reporter gene, Nos terminator and ampicillin resistance selection marker.

FIG. 57 is nucleotide sequence of *Emp1* from Litts, J.C. et al., Plant Mol. Biol. 19: 335-337 (1992).

FIG. 58 shows a histochemical analysis of transgenic line 257-10-1 and tissue-specific expression of *Emp1*/GUS in transgenic rice. Panel a is a section of a mature seed. Panel b is a section of a mature embryo. Panel c is a leaf section. Panel d is a

root section. Panel e is a stem cross-section. Panel f is a section of a non-transgenic rice seed of Taipei 309 as control.

FIG. 59 is a dark-field photomicrograph of mature seed from transgenic line 257-10-1 showing restricted cellular expression of Emp1/GUS to aleurone and embryo. Panel a is a view of an embryo. Panel b is a view of a thick layer aleurone. Panel c is a view of a thin layer aleurone. Panel d is a view of Taipei 309 negative control embryo. Panel e is a view of Taipei 309 negative control thick layer aleurone. Panel f is a view of Taipei 309 negative control thin layer aleurone.

Detailed Description of the Invention

I. Definitions

Unless otherwise indicated, all terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook *et al.* (1989) Molecular Cloning: A Laboratory Manual (Second Edition), Cold Spring Harbor Press, Plainview, N.Y. and Ausubel FM *et al.* (1993) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y., for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary. For a better understanding of the present invention, the terms used herein has the following particular meanings:

As used herein, the term "active" refers to any biological activity associated with a particular polypeptide, whether directly or indirectly, and includes but is not limited to therapeutic, anti-infective, prophylactic or diagnostic activities. Such activity may be, for example, an enzymatic activity associated with human lysozyme.

As used herein, the term "anti-bacterial polypeptide" refers to a polypeptide that has bacteriostatic or bactericidal activity.

As used herein, the term "anti-microbial polypeptide" refers to a polypeptide that has anti-microbial activity, including but not limited to anti-bacterial, antiviral, anti-fungal and anti-parasitic activities and can include acute phase proteins, cationic anti-microbial peptides and probiotic proteins. Such anti-microbial polypeptides are capable of inhibiting the growth of one or more of Gram-negative bacteria, Gram-positive bacteria, fungi (including yeast), parasites (including planaria and nematodes) and viruses. Typically, such anti-microbial peptides exhibit selective biological activity against such microbes over eukaryotic cells.

As used herein, the term "Bar gene" refers to a nucleotide sequence encoding a phosphinothricin acetyltransferase enzyme that upon expression confers resistance to the herbicide glufosinate-ammonium ("Basta").

"Cereal adjuncts" as used herein refers to cereal grains, principally barley, wheat, rye, oats, maize, sorghum and rice, or processed whole or portions thereof, especially the starch fraction, which are added to the barley mash, which allows the barley enzymes to hydrolyze both the barley starch and the starch derived from the cereal adjunct. "Transgenic cereal adjuncts" as used herein refers to transgenic cereal grains, principally barley, wheat, rye, oats, maize, sorghum and rice, and which is expressing a recombinant molecule in a grain part, principally the endosperm (starch) layer.

"Conversion" as used herein refers to the process of starch hydrolysis, usually catalyzed by acid or enzyme action, which produces dextrose, maltose, and higher polysaccharides from starch.

Chimeric gene" as defined herein refers to a nucleic acid construct for introduction into a host and may include a combination of parts of different genes of exogenous or autologous origin, including regulatory elements.

As used herein, the term "expression" with respect to a polypeptide refers to a process by which the polypeptide is produced in a host based on the nucleic acid sequence of a gene. The process includes both transcription and translation. The term "expression" may also be used with respect to the generation of RNA from a DNA sequence.

An "expression cassette" is a nucleic acid construct generated recombinantly or synthetically, that contains a series of specified nucleic acid elements that can be transcribed or translated to produce one or more recombinant polypeptides in a host expression system. The expression cassette can be incorporated into a plasmid or a viral vector, for example, to form an expression vector, or can be integrated into host chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment, for example, by particle bombardment. Typically, the expression cassette includes, among other sequences, a promoter, a transcription start, a translation start, a heterologous gene of interest, a translation terminator and a transcription terminator. Optionally, the expression cassette may contain one or more selectable markers.

An "expression vector" refers to a vector that contains or is suitable for use with an expression cassette for expression of heterologous DNA or RNA in a host cell. Many prokaryotic and eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill

in the art. Optionally, an expression vector may contain one or more selectable markers for selection of hosts that contain the expression vector.

As used herein, the term "gene" means a DNA or RNA fragment, including genomic or complementary sequences, that encodes a polypeptide, and may or may not include regions preceding and following the coding region, such as 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons).

The term "grain," as used herein, refers to the fruit of a mature seed-producing plant and includes the seed. In the context of the present invention, the term "grain" and "seed" is used interchangeably.

The term "introduced" in the context of inserting a nucleic acid sequence into a cell, includes biological or mechanical means of introduction such as transfection, transformation, transduction, electroporation, and particle bombardment. Such introduced sequences may be incorporated into a host cell genome such as into the chromosome, plasmid, plastid, or mitochondrial DNA for stable expression, or in the form of autonomous replicon, that may be transiently expressed (for example, transfected mRNA).

"Heterologous nucleic acid construct" refers to a nucleic acid molecule, typically for introduction into a host cell for expression purposes, that is not native to the host, although parts of the construct may be native host sequences.

By "host cell" is meant a cell which contains an introduced vector, plasmid or heterologous nucleic acid construct and supports the replication, transcription and/or translation, that is, expression thereof. Host cells can be prokaryotic cells, such as *E. coli*, or eukaryotic cells such as yeast or other fungi, plant, insect, amphibian, or mammalian cells. Plant host cells can be monocotyledenous or dicotyledenous plant cells, whether in tissues such as roots or leaves or in grains or seeds.

"Malt" as used herein is a dried grain resulting from limited germination of cereal grains, usually barley, under controlled conditions, but other cereals can be malted as well. When malted, the grain produces relatively large amounts of α- and β-amylases, which are important in the conversion of starch to dextrins and fermentable sugars.

The basic "Malting process" consists of three steps: steeping; germination; and drying, as described in Hickenbottom, J.W., Cereal Foods World, 41(10): 788-710 (1996) and QIB Technical Bulletin, Vol. V (3):1-8 (1983). The germinated malt is called a "green malt." Drying or kilning of the green malt stops the germination process. The end result of the malting process is a "dry malted grain" or malted barley.

A "Liquid malt extract" results from the dry malted grain being coarsely ground, water extracted and concentrated. Many enzymatic systems, such as α- and β-amylases and carbohydrase, are activated during germination. After addition of water and during a series of time and temperature changes, the enzyme systems in the malted grain break down the endosperm cell structure and some of the starch is converted into fermentable sugars by these enzymes. If cereal adjuncts, usually corn grits, are added at this time to provide additional starch, the resulting malt syrups will be mellower and sweeter in flavor than the extracts. After the appropriate conditions are met, such as the appropriate extent of hydrolytic activity or liquefaction, the liquid extract, now a "wort" can be moved to an evaporator for conversion to a "malt syrup." The remaining grain residue, separated from the liquid extract, can be used as a high protein feed, such as for animal feed.

"Malt extract" refers to viscous concentrates of water extract of dried malt. "Dry malt extracts" are produced by spray drying the liquid malt extracts.

The term "milk protein" or "milk polypeptide" refers to polypeptides that are present in human breast milk or milk of non-human animals, including lactoferrin, lysozyme, lactohedrin, haptocorrin, epidermal growth factor, insulin-like growth factor I, kappa-casein, immunoglobulins, albumin, lactoglobulin, lactoperoxidase, alpha-1-antitrypsin, for example.

As used herein, the terms "native" relative to a given cell, polypeptide, nucleic acid, trait or phenotype refers to the form in which that is typically found in nature.

A nucleic acid is "operably linked" when it is placed into a functional relationship with other nucleic acid sequences. For example, DNA for a leader sequence, such as a signal peptide, is operably linked to heterologous gene if the combination is expressed as a preprotein that can be cleaved or directed to the appropriate cell compartment for cleavage to generate a desired heterologous polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a leader sequence, contiguous and in reading phase. However, some "operably linked' elements, *e.g.*, enhancers, do not have to be contiguous. Linking may be accomplished by ligation at convenient restriction sites. If such sites do not exist, synthetic oligonucleotide adaptors or linkers may be used in accordance with conventional practice.

The term "Percent or % homology" is used interchangeably herein with the term "% identity" and refers to the level of nucleic acid or amino acid sequence identity between two or more aligned sequences, when aligned using a sequence alignment program. For example, 70% homology means the same thing as 70% sequence identity determined by a defined algorithm. For purposes herein, a homologue of a given sequence has greater than 80% sequence identity over a length of the given sequence and include, but are not limited to, 80%, 85%, 90% or 95% or more percent sequence identity to a given sequence, such as, the coding sequence for lactoferrin, as described herein.

As used herein, a "plant cell" refers to any cell derived from a plant, including undifferentiated tissue (*e.g.*, callus) as well as plant seeds, pollen, progagules and embryos.

As used herein, the term "plant" includes reference to whole plants, plant tissues and individual plant cells, and progeny of same. Thus, the term includes, without limitation, leaves, stems, roots, shoots, endosperms, grains, seeds, embryos, suspension cultures, meristematic regions, callus tissue, gametophytes, sporophytes, pollen, progagules, and microspores. The class of plants includes higher plants amenable to transformation techniques, such as monocotyledenous and dicotyledenous plants.

As used herein, a "plant cell" refers to any cell derived from a plant.

As used herein, the term "plasmid" refers to a nucleic acid construct, typically circular, double-stranded or single stranded, containing DNA or RNA, used as a vector for introduction of foreign or heterologous nucleic acid into a host cell, and which forms an extrachromosomal self-replicating genetic element in host cells, such as bacteria or eukaryotic cells.

The term "polypeptide" as used herein refers to a compound made up of a sequence of amino acid residues linked by peptide bonds and includes peptides, oligopeptides, proteins, and active fragments thereof. The term "protein" as used herein may be synonymous with the term "polypeptide" or may refer, in addition, to a complex of two or more polypeptides and may be in primary, secondary or tertiary configuration.

As used herein, the term "promoter" refers to a nucleic acid sequence functioning to direct transcription of a downstream gene. The promoter will generally be appropriate to the host cell in which the target gene is being expressed. The promoter, together with other transcriptional and translational regulatory nucleic acid sequences are termed "control sequences". Thus, for example, transcriptional and translational

regulatory sequences include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. A promoter may be a constitutive promoter or inducible, for example, during germination of a seed or during seed maturation.

As used herein, the term "purifying" is used interchangeably with the term "isolating" and generally refers to any separation of a particular component from other components of the environment in which it is found or produced. For example, purifying a recombinant protein from plant cells in which it was produced typically means subjecting transgenic protein-containing plant material to separation techniques such as sedimentation, centrifugation, filtration, column chromatography. The results of any of such purifying or isolating steps may still contain other components as long as the results have less other components ("contaminating components") than before such purifying or isolating steps.

As used herein, "recombinant" includes reference to a molecule such as nucleic acid or polypeptide or to a host cell containing such, where the molecules are produced by introduction into and expression of an introduced nucleic acid sequence in the host cell. Thus, for example, recombinant cells express genes that are not found in identical form within the native cell or express native genes that are not otherwise expressed to such a level in the native cell without deliberate human intervention.

As used herein, the term "seed" refers to all seed components, including, for example, the coleoptile and leaves, radicle and coleorhiza, scutulum, starchy endosperm, aleurone layer, pericarp and/or testa, either during seed maturation and seed germination. In the context of the present invention, the term "seed" and "grain" is used interchangeably.

As used herein "seed in a form for use as a food or food supplement" includes, but is not limited to, processed seed fractions such as de-hulled whole seed, flour (seed that has been de-hulled by milling and grounding into a powder form) a seed protein extract (where the protein fraction of the flour has been separated from the carbohydrate fraction) and/or a purified protein fraction derived from the transgenic grain.

As used herein, the term "selectable marker-encoding nucleotide sequence" or "selectable marker" refers to a nucleotide sequence or the product thereof that confers to a host cell to which it is introduced the ability to grow in the presence of a selective

agent, such as antibiotic resistance, including resistance to ampicillin and hygromycin, for example.

As used herein, the term "sequence identity" refers to the degree of similarity or identity between two or more nucleic acid or amino acid sequences aligned using a sequence alignment program. Exemplary computer programs which can be used to determine identity between two sequences include, but are not limited to, the suite of BLAST programs, *e.g.*, BLASTN, BLASTX, and TBLASTX, BLASTP and TBLASTN, publicly available on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/". See, also, Altschul, S.F. *et al.*, 1990 and Altschul, S.F. *et al.*, 1997.

Sequence searches are typically carried out using the BLASTN program when evaluating a given nucleic acid sequence relative to nucleic acid sequences in the GenBank DNA Sequences and other public databases. The BLASTX program is preferred for searching nucleic acid sequences which have been translated in all reading frames against amino acid sequences in the GenBank Protein Sequences and other public databases. Both BLASTN and BLASTX are run using default parameters of an open gap penalty of 11.0, and an extended gap penalty of 1.0, and utilize the BLOSUM-62 matrix. [See, Altschul, *et al.*, 1997.]

A preferred alignment of selected sequences in order to determine "% identity" between two or more sequences, is performed using for example, the CLUSTAL-W program in MacVector version 6.5, operated with default parameters, including an open gap penalty of 10.0, an extended gap penalty of 0.1, and a BLOSUM 30 similarity matrix.

A nucleic acid sequence is considered to be "selectively hybridizable" to a reference nucleic acid sequence if the two sequences specifically hybridize to one another under moderate to high stringency hybridization and wash conditions. Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex or probe. For example, "maximum stringency" typically occurs at about Tm-5°C (5° below the Tm of the probe); "high stringency" at about 5-10° below the Tm; "intermediate stringency" at about 10-20° below the Tm of the probe; and "low stringency" at about 20-25° below the Tm. Functionally, maximum stringency conditions may be used to identify sequences having strict identity or near-strict identity with the hybridization probe; while high stringency conditions are used to identify sequences having about 80% or more sequence identity with the probe.

Moderate and high stringency hybridization conditions are well known in the art (see, for example, Sambrook *et al,* 1989, Chapters 9 and 11, and in Ausubel *et al.*,

1993, expressly incorporated by reference herein). An example of high stringency conditions includes hybridization at about 42°C in 50% formamide, 5X SSC, 5X Denhardt's solution, 0.5% SDS and 100 µg/ml denatured carrier DNA followed by washing two times in 2X SSC and 0.5% SDS at room temperature and two additional times in 0.1X SSC and 0.5% SDS at 42°C.

As used herein, the terms "transformed" or "transgenic" with reference to a host cell means the host cell contains a non-native or heterologous or introduced nucleic acid sequence that is absent from the native host cell. Further, "stably transformed" in the context of the present invention means that the introduced nucleic acid sequence is maintained through two or more generations of the host, which may be due to integration of the introduced sequence into the host genome.

"Transgenic malt extract" as used herein refers to a vicious concentrate of the water extract of dried malt which includes a recombinant polypeptide and/or metabolite.

"Transgenic malt syrup" as used herein refers to vicious concentrate of the water extract of dried 'malt' and other cereal grains which includes a recombinant polypeptide and/or metabolite.

As used herein, the term "transgenic plant" refers to a plant that contains introduced nucleic acid sequences that are not present in the native ("untransformed") plant or plant cell. Thus a plant having within its cells a heterologous polynucleotide is referred to herein as a "transgenic plant." The heterologous polynucleotide can be either stably integrated into the genome, or can be extra-chromosomal. Preferably, the polynucleotide of the present invention is stably integrated into the genome such that the polynucleotide is passed on to successive generations. The term "transgenic" as used herein does not encompass the alteration of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation. "Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of the introduced nucleic acids including those transgenics initially so altered as well as those created by sexual crosses or asexual reproduction of the initial transgenics. Thus, a transgenic or transformed plant cell or plant also includes progeny of such cell or plant and progeny produced from a breeding program employing one or more transgenic plants as parents in one or more crosses and exhibiting an altered phenotype resulting from the presence of the introduced sequence. Hence, a transformed or transgenic plant herein includes

any plant which has a cell containing introduced nucleic acid sequences, regardless of whether the sequence was introduced into the plant directly through transformation or by generational transfer from a progenitor cell which originally received the construct by direct transformation.

As used herein, the term "vector" refers to a nucleic acid construct designed for introduction of a heterologous nucleic acid into host cells.

II. Expression Systems, Methods and Compositions of the Present Invention

The present invention provides for nucleic acid constructs, vectors, expression systems and methods for high level expression of recombinant polypeptides in plants, including monocots and dicots, and compositions containing such as well as compositions resulting from such expression. For example, the monocots are cereals including rice, wheat, barley, oats, millet, sorghum and corn; and the dicots are plants such as tobacco or cotton. Some of the same or similar nucleic acid constructs as made for expression of heterologous polypeptides in monocots can be used for expression in dicots as well. Preferably, monocot regulatory sequences are used in monocots and dicot regulatory sequences are used in dicots. Further genes to be expressed in monocots are codon-optimized for expression in monocots, and genes to be expressed in dicots are codon-optimized for expression in dicots.

In practicing the invention, a heterologous polypeptide is produced in the seeds or grain of transgenic plants which express the nucleic acid coding sequence for the heterologous polypeptide. The transgenic grains may thereafter be orally delivered, such as, for example, as a tablet or capsule or an extract or malt syrup for use as a food additive or supplement to produce an "improved food compositions".

More specifically, the present invention is based on the expression of nucleic acid molecules encoding heterologous polypeptides which are each linked to a signal peptide for directing the expressed polypeptide to the endosperm, under the control of one or more seed or maturation specific promoters, such as a promoter from a seed storage protein, with or without the addition of one or more transcription factors.

The expression of heterologous polypeptides in plants pursuant to the present invention is exemplified herein in detail by the expression of recombinant codon optimized human lactoferrin (hLF), human lysozyme (hLys) and α_1-antitrypsin. Plasmids suitable for expression of other recombinant polypeptides are also exemplified. It is understood that other polypeptides can be expressed in a similar fashion by substituting

the gene in the present expression cassettes with the gene of interest, with or without codon optimization.

The present invention provides food supplement compositions (also termed "improved food compositions" comprising heterologous polypeptides and methods of making such compositions. In practicing the invention, a heterologous polypeptide is produced in the seeds or grain of transgenic plants which express the nucleic acid coding sequences for the heterologous polypeptides and the transgenic grains added to a food such as an infant formula, bakery goods, etc., to result in an "improved food compositions". More specifically, the invention is based on the expression of heterologous polypeptides in grains or seeds of plants, such as monocot plants, as exemplified herein by the expression of human lactoferrin (hLF), human lysozyme and human α_1-antitrypsin, under the control of a seed specific promoter in rice. The heterologous polypeptides produced by transgenic plants are compared to the native form of the same protein, information on the stability of the recombinant protein and the advantages of using rice, cereal or other grain containing such recombinant heterologous polypeptides in foods and in industrial applications, are further described below.

The invention relies on the use of heterologous nucleic acid constructs including the coding sequences for commercially important heterologous polypeptides of nutritional, therapeutic and/or industrial value, as exemplified herein by lactoferrin and lysozyme.

The improved food compositions of the invention include milk proteins such as lactoferrin, and lysozyme, produced in the grain of transgenic plants, which are useful for improved nutrition. In one preferred approach, the improved food compositions are administered to an infant. Typically the improved food compositions, *e.g.*, infant formula contain one or more recombinant human milk proteins in an amount that corresponds to the amount and proportions of the same human milk proteins found in endogenous human milk.

A. Lysozyme

Human milk lysozyme, called muramidase or peptidoglycan N-acetylmuramoyl-hydrolase (EC 3.2.1.17) contains 130 amino acid residues and is a protein of 14.7 kDa in size. Human lysozyme is non-glycosylated and possesses unusual stability *in vitro* and *in vivo* due to its amino acid and secondary structure.

Lysozymes act as enzymes that cleave peptidoglycans, and ubiquitous cell wall component of microorganisms, in particular bacteria. Specifically, lysozymes are 1,4-acetylmuramidases that hydrolyze the glycoside bond between N-acetylmuramic acid and N-acetylglucosamine. Gram-positive bacteria are highly susceptible to lysozyme due to the polypeptidoglycan on the outside of the cell wall. Gram-negative strains have a single polypeptidoglycan layer covered by lipopolysaccharides and are therefore less susceptible to lysis by lysozyme, however, the sensitivity can be increased by the addition of EDTA (Schütte and Kula, 1990).

The recombinant human lysozyme expressed according to the present invention was tested for its functional activity using an anti-microbial assay developed and described herein.

B. Lactoferrin

Lactoferrin is an iron-binding protein found in the granules of neutrophils where it apparently exerts an antimicrobial activity by withholding iron from ingested bacteria and fungi; it also occurs in many secretions and exudates (milk, tears, mucus, saliva, bile, etc.). In addition to its role in iron transport, lactoferrin has bacteriostatic and bactericidal activities, in addition to playing a role as an anti-oxidant (Satue-Gracia *et al.*, 2000).

The mature lactoferrin (LF) polypeptide consists of 692 amino acids, consists of a single-chain polypeptide that is relatively resistant to proteolysis, is glycosylated at two sites (N138 and N478) and has a molecular weight of about 80 kD.

C. Lactoperoxidase is an enzyme which catalyzes the conversion of hydrogen peroxide to water. This enzyme is found in human milk, and plays host defensive roles through antimicrobial activity. When hydrogen peroxide and thiocyanate are added to raw milk, the SCN's oxidized by the enzyme-hydrogen peroxide complex producing bactericidal compounds which destroy Gram-negative bacteria (Shin).

D. Kappa-Casein. This group of proteins are readily digested and account for almost half of the protein content in human milk They are important as nutritional protein for breast-fed infants. It has also been advocated that part of the antimicrobial activity of human milk resides in the caseins, most likely the glycosylated kappa-casein.

E. Alpha-1-antitrypsin ("AAT"). AAT belongs to the class of serpin inhibitors, has a molecular mass of 52 kD, and contains about 15% carbohydrate (Carrell et al, 1983).

While the binding affinity of AAT is highest for human neutrophil elastase, it also has affinity for pancreatic proteases such as chymotrypsin and trypsin (Beatty et al., 1980).

F. Lactadherin is a protective glycoprotein present in human milk that helps protect breast-fed infants against infection by microorganisms. Protection against certain virus infections by human milk is also associated with lactadherin. (Newburg, 1999, 1998; Peterson; Hamosh).

G. Epidermal Growth Factor and Insulin-like Growth Factor-1 are two growth factors present in human milk. These molecules may stimulate growth and development of the infant gastrointestinal tract. (Murphy; Prosser).

H. Immunoglobulins present in human act to confer resistance to a variety of pathogens to which the mother may have been exposed. (See, for example, Humphreys; Kortt; Larrick; Maynard; and Peeters).

III. Expression Vectors For Generation Of Transgenic Plants Expressing Human Milk Proteins

Expression vectors for use in the present invention are chimeric nucleic acid constructs (or expression vectors or cassettes), designed for operation in plants, with associated upstream and downstream sequences.

In general, expression vectors for use in practicing the invention include the following operably linked components that constitute a chimeric gene: (i) a seed maturation-specific promoter from a plant, (ii) operably linked to a leader DNA encoding a monocot seed-specific transit sequence capable of targeting a linked polypeptide to a seed of the plant, such as the leader sequence for targeting to a protein-storage body, and (iii) a heterologous polypeptide-encoding sequence.

The chimeric gene, in turn, is typically placed in a suitable plant-transformation ("expression") vector having (i) companion sequences upstream and/or downstream of the chimeric gene which are of plasmid or viral origin and provide necessary characteristics to the vector to permit the vector to move DNA from one host to another, such as from bacteria to a desired plant host; (ii) a selectable marker sequence; and (iii) a transcriptional termination region with or without a polyA tail.

Exemplary methods for constructing chimeric genes and transformation vectors carrying the chimeric genes are given in the examples below.

A. Promoters

In one aspect of this embodiment, the expression construct includes promoters from genes that are upregulated activity during seed maturation. Examples of such promoters include, but are not limited to the maturation-specific promoter associated with one of the following maturation-specific monocot storage proteins: rice glutelins, oryzins, and prolamines, barley hordeins, wheat gliadins and glutenins, maize zeins and glutelins, oat glutelins, and sorghum kafirins, millet pennisetins, rye secalins. Also included herein are aleurone and embryo specific promoters associated with the rice, wheat and barley genes such as lipid transfer protein Ltp1, chitinase Chi26 (Hwang, Y.S. et al., Plant Cell Rep. 20: 647-654 (2001)), and Em protein Emp1 (Litts, J.C., et al., Plant Mol. Biol. 19: 335-337 (1992)). Exemplary regulatory regions from these genes are exemplified by SEQ ID NOS: 15-23, as identified in the Description of the Sequences.

In one embodiment of the present invention, a heterologous nucleic acid encoding a heterologous polypeptide is expressed under the control of a promoter from a transcription initiation region that is preferentially expressed in plant seed tissue. Examples of such seed preferential transcription initiation sequences include those derived from sequences encoding plant storage protein genes or from genes involved in fatty acid biosynthesis in oilseeds. Exemplary preferred promoters include a glutelin (Gt-1) promoter, as exemplified by SEQ ID NO: 18, which effects gene expression in the outer layer of the endosperm and a globulin (Glb) promoter, as exemplfieid by SEQ ID NO: 16, which effects gene expression in the center of the endosperm. Promoter sequences for regulating transcription of gene coding sequences operably linked thereto include naturally-occurring promoters, or regions thereof capable of directing seed-specific transcription, and hybrid promoters, which combine elements of more than one promoter. Methods for construction such hybrid promoters are well known in the art.

In some cases, the promoter is derived from the same plant species as the plant cells into which the chimeric nucleic acid construct is to be introduced. Promoters for use in the invention are typically derived from cereals such as rice, barley, wheat, oat, rye, corn, millet, triticale or sorghum.

Alternatively, a seed-specific promoter from one type of plant, such as a monocot, may be used to regulate transcription of a nucleic acid coding sequence from a different plant, whether a cereal monocot, a non-cereal monocot or a dicot.

Numerous types of appropriate expression vectors, and suitable regulatory sequences are known in the art for a variety of plant host cells. The transcription

regulatory or promoter region is chosen to be regulated in a manner allowing for induction under seed-maturation conditions. Other promoters suitable for expression in maturing seeds include the barley endosperm-specific B1-hordein promoter (Brandt, A., et al., (1985), Glub-2 promoter, Bx7 promoter, Gt3 promoter, Glub-1 promoter and Rp-6 promoter, particularly if these promoters are used in conjunction with transcription factors. The primary structure of a B1 hordein gene from barley is provided in Carlsberg Res. Commun. 50, 333-345).

B. Signal/targeting/transport Sequences

In addition to encoding the protein of interest, the expression cassette or heterologous nucleic acid construct may encode a signal/targeting/transport peptide that allows processing and translocation of the protein, as appropriate. Exemplary signal/targeting/transport sequences, particularly for targeting proteins to intracellular bodies, such as vacuoles, are signal/targeting sequences associated with the monocot maturation-specific genes: glutelins, prolamines, hordeins, gliadins, glutenins, zeins, albumin, globulin, ADP glucose pyrophosphorylase, starch synthase, branching enzyme, Em, and lea. Exemplary sequences encoding a leader sequence for protein storage body are identified herein as SEQ ID NOS: 24-30.

In one embodiment of the present invention, the method is directed toward the localization of heterologous polypeptide expression in a given subcellular compartment or tissue, such as protein-storage body, aleurone layers or embryo, but also including other compartments such as vacuoles, chloroplasts or other plastidic compartments. For example, when heterologous polypeptide expressed is targeted to plastids, such as chloroplasts, the construct employs the use of sequences to direct the gene to the plastid. Such sequences are for example chloroplast transit peptides (CTP) or plastid transit peptides (PTP). In this manner, when the gene of interest is not directly inserted into the plastid, the expression construct additionally contains a gene encoding a transit peptide to direct the gene of interest to the plastid. The chloroplast transit peptides may be derived from the gene of interest, or may be derived from a heterologous sequence having a CTP. Such transit peptides are known in the art. See, for example, Von Heijne *et al.,* 1991; Clark *et al.,* 1989; della-Cioppa *et al.,* 1987; Romer *et al.,* 1993; and Shah *et al.*, 1986. Additional transit peptides for the translocation of the protein to the endoplasmic reticulum (ER) (Chrispeels, K., 1991), nuclear localization signals (Raikhel, 1992), or vacuole may also find use in the constructs of the present invention.

Another exemplary class of signal/targeting/transport sequences are sequences effective to promote secretion of heterologous protein from aleurone cells during seed germination, including the signal sequences associated with α-amylase, protease, carboxypeptidase, endoprotease, ribonuclease, DNase/RNase, (1-3)- β-glucanase, (1-3)(1-4)- β-glucanase, esterase, acid phosphatase, pentosamine, endoxylanase, *β*-xylopyranosidase, arabinofuranosidase, β-glucosidase, (1-6)- β-glucanase, perioxidase, and lysophospholipase.

Since many protein storage proteins are under the control of a maturation-specific promoter, and this promoter is operably linked to a leader sequence for targeting to a protein body, the promoter and leader sequence can be isolated from a single protein-storage gene, then operably linked to a heterologous polypeptide in a chimeric gene construct. One exemplary promoter-leader sequence is from the rice *Gt1* gene, having an exemplary sequence identified by SEQ ID NO:15. Alternatively, the promoter and leader sequence may be derived from different genes. Another exemplary promoter/leader sequence combination is the rice *Glb* promoter linked to the rice *Gt1* leader sequence, as exemplified by SEQ ID NO: 16.

C. Codon Optimization

It has been shown that production of recombinant protein in transgenic barley grain was enhanced by codon optimization of the gene (Horvath *et al.*, 2000; Jensen *et al.*, 1996). The intent of codon optimization was to change an A or T at the third position of the codons of G or C. This arrangement conforms more closely with codon usage in typical rice genes (Huang *et al.*, 1990a).

To obtain a high expression level for heterologous polypeptides in rice cells in the examples herein, the polypeptide coding sequences used were codon optimized. The G + C content was thus increased from 46% to 68%. The codon-optimized lysozyme coding sequence for use in practicing the invention, for example, is presented as SEQ ID NO:1.

Similarly, in another example, for high level expression of human lactoferrin (hLF) in rice cells, the native hLF coding sequence was codon optimized. Out of 693 codons used in the lactoferrin gene, 413 codons were changed by one or two nucleotides. The amino acid sequence of LF was unchanged. The codon optimized lactoferrin coding sequence for use in practicing the invention is presented as SEQ ID NO:3.

Codon optimized sequences for heterologous polypeptides exemplified herein are given as follows: for lactoferricin, SEQ ID NO: 7; for EGF, SEQ ID NO: 8; for IGF-1, SEQ ID NO: 9; for lactohedrin, SEQ ID NO: 10; for kappa-casein, SEQ ID NO: 11; for

haptocorrin, SEQ ID NO: 12; for lactoperoxidase, SEQ ID NO: 13; and for alpha-1-antitrypsin, SEQ ID NO: 14.

D. Transcription Factors

In one embodiment of the invention, the transgenic plant herein is also transformed with the coding sequence of one or more transcription factors capable of enhancing the expression of a maturation-specific promoter. For example, the embodiment involves the use of the maize *Opaque 2* (O2) or prolamin box binding factor (PBF), separately or together, or the use of rice endosperm bZip (Reb) protein as transcriptional activators herein. Exemplary sequence for these three transcription factors are given identified below as SEQ ID NOS: 31-33. Transcription factor sequences and constructs applicable to the present invention are detailed in co-owned PCT application No. PCT/US01/14234, International Publication number WO 01/83792 A1, published November 8, 2001, which is incorporated herein by reference in its entireties, including the references cited therein. Reference is also made to Examples 12-15 herein.

Transcription factors are capable of sequence-specific interaction with a gene sequence or gene regulatory sequence. The interaction may be direct sequence-specific binding in that the transcription factor directly contacts the gene or gene regulatory sequence or indirect sequence-specific binding mediated by interaction of the transcription factor with other proteins. In some cases, the binding and/or effect of a transcription factor is influenced (in an additive, synergistic or inhibitory manner) by another transcription factor. The gene or gene regulatory region and transcription factor may be derived from the same type (*e.g.*, species or genus) of plant or a different type of plant. The binding of a transcription factor to a gene sequence or gene regulatory sequence may be evaluated by a number of assays routinely employed by those of skill in the art, for example, sequence-specific binding may be evaluated directly using a label or through gel shift analysis.

The ability of a given transcription factor to enhance the expression of a heterologous polypeptide under a given promoter may vary. The present invention includes use of the following exemplary combinations of transcription factors and promoters: Reb-Glb promoter; O2-Glb promoter; PBF-Glb; O2-PBF-Glb; O2-RP6; PBF-RP6; O2-PBF-RP6; O2- PG5a; PBF-PG5a; O2-PBF-PG5a; O2- Bx7; PBF-Bx7; O2-PBF-Bx7; O2-Gt1; PBF-Gt1; O2-PBF-Gt1; all with or without signal peptide sequences from protein body genes.

As detailed in the cited PCT application, the transcription factor gene is introduced into the plant in a chimeric gene containing a suitable promoter, preferably a maturation-specific seed promoter operably linked to the transcription factor gene. Plants may be stably transformed with a chimeric gene containing the transcription factor by methods similar to those described with respect to the milk-protein gene(s), exemplified herein. Plants stably transformed with both exogenous transcription factor(s) and heterologous polypeptide genes may be prepared by co-transforming plant cells or tissue with both gene constructs, selecting plant cells or tissue that have been co-transformed, and regenerating the transformed cells or tissue into plants. Alternatively, different plants may be separately transformed with exogenous transcription factor genes and milk-protein genes, then crossed to produce plant hybrids containing by added genes.

E Additional Expression Vector Components

Expression vectors or heterologous nucleic acid constructs designed for operation in plants, comprise companion sequences upstream and downstream to the expression cassette. The companion sequences are of plasmid or viral origin and provide necessary characteristics to the vector to permit the vector to move DNA from one host to another such as from bacteria to the plant host including, for example, sequences containing an origin of replication and a selectable marker. Typical secondary hosts for production of plasmids for transformation into plants include bacteria and yeast.

In one embodiment, the secondary host is *E. coli*, the origin of replication is a colE1-type, and the selectable marker is a gene encoding ampicillin resistance. Such sequences are well known in the art and are commercially available as well (*e.g.*, Clontech, Palo Alto, Calif.; Stratagene, La Jolla, CA).

The transcription termination region may be taken from a gene where it is normally associated with the transcriptional initiation region or may be taken from a different gene. Exemplary transcriptional termination regions include the NOS terminator from *Agrobacterium* Ti plasmid and the rice α-amylase terminator.

Polyadenylation tails (Alber *et al.*, 1982) may also be added to the expression cassette to optimize high levels of transcription and proper transcription termination, respectively. Polyadenylation sequences include, but are not limited to, the *Agrobacterium* octopine synthetase signal, Gielen, *et al.*, 1984 or the nopaline synthase of the same species, Depicker, *et al.*, 1982.

Suitable selectable markers for selection in plant cells include, but are not limited to, antibiotic resistance genes, such as, kanamycin (*npt*II), G418, bleomycin, hygromycin, chloramphenicol, ampicillin, tetracycline, and the like. Additional selectable markers include a bar gene which codes for bialaphos resistance; a mutant EPSP synthase gene which encodes glyphosate resistance; a nitrilase gene which confers resistance to bromoxynil; a mutant acetolactate synthase gene (ALS) which confers imidazolinone or sulphonylurea resistance; and a methotrexate resistant DHFR gene.

The particular marker gene employed is one which allows for selection of transformed cells as compared to cells lacking the DNA which has been introduced. Preferably, the selectable marker gene is one which facilitates selection at the tissue culture stage, *e.g.*, a kanamyacin, hygromycin or ampicillin resistance gene.

The vectors of the present invention may also be modified to include intermediate plant transformation plasmids that contain a region of homology to an *Agrobacterium tumefaciens* vector, a T-DNA border region from *Agrobacterium tumefaciens*, and chimeric genes or expression cassettes (described above). Further, the vectors of the invention may comprise a disarmed plant tumor inducing plasmid of *Agrobacterium tumefaciens*.

In general, a selected nucleic acid sequence is inserted into an appropriate restriction endonuclease site or sites in the vector. Standard methods for cutting, ligating and *E. coli* transformation, known to those of skill in the art, are used in constructing vectors for use in the present invention. (See generally, Maniatis, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2d Edition (1989); Ausubel, *et al.*, (c) 1987, 1988, 1989, 1990, 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, N.Y; and Gelvin, S.B., *et al.*, eds. PLANT MOLECULAR BIOLOGY MANUAL, (1990), all three of which are expressly incorporated by reference, herein.

IV. Generation of Transgenic Plants

Plant cells or tissues are transformed with expression constructs (heterologous nucleic acid constructs), for example, plasmid DNA, into which the gene of interest has been inserted) using a variety of standard techniques. It is preferred that the vector sequences be stably integrated into the host genome.

The method used for transformation of host plant cells is not critical to the present invention. For commercialization of recombinant heterologous polypeptides expressed in accordance with the present invention, the transformation of the plant is preferably permanent, *i.e.* by integration of the introduced expression constructs into the

host plant genome, so that the introduced constructs are passed onto successive plant generations. The skilled artisan will recognize that a wide variety of transformation techniques exist in the art, and new techniques are continually becoming available.

Any technique that is suitable for the target host plant may be employed within the scope of the present invention. For example, the constructs can be introduced in a variety of forms including, but not limited to, as a strand of DNA, in a plasmid, or in an artificial chromosome. The introduction of the constructs into the target plant cells can be accomplished by a variety of techniques, including, but not limited to calcium-phosphate-DNA co-precipitation, electroporation, microinjection, *Agrobacterium*-mediated transformation, liposome-mediated transformation, protoplast fusion or microprojectile bombardment. The skilled artisan can refer to the literature for details and select suitable techniques for use in the methods of the present invention. Exemplary methods for plant transformation are given in Example 2.

When *Agrobacterium* is used for plant cell transformation, a vector is introduced into the *Agrobacterium* host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the *Agrobacterium* host. The Ti- or Ri-plasmid containing the T-DNA for recombination may be armed (capable of causing gall formation) or disarmed (incapable of causing gall formation), the latter being permissible, so long as the *vir* genes are present in the transformed *Agrobacterium* host. The armed plasmid can give a mixture of normal plant cells and gall.

In some instances where *Agrobacterium* is used as the vehicle for transforming host plant cells, the expression or transcription construct bordered by the T-DNA border region(s) is inserted into a broad host range vector capable of replication in *E. coli* and *Agrobacterium*, examples of which are described in the literature, for example pRK2 or derivatives thereof. See, for example, Ditta *et al.*, 1980 and EPA 0 120 515, expressly incorporated by reference herein. Alternatively, one may insert the sequences to be expressed in plant cells into a vector containing separate replication sequences, one of which stabilizes the vector in *E. coli*, and the other in *Agrobacterium* See, for example, McBride *et al.*, 1990, wherein the pRiHRI (Jouanin, *et al., 1985,* origin of replication is utilized and provides for added stability of the plant expression vectors in host *Agrobacterium* cells.

Included with the expression construct and the T-DNA is one or more selectable marker coding sequences which allow for selection of transformed Agrobacterium and transformed plant cells. A number of markers have been developed for use with plant cells, such as resistance to chloramphenicol, kanamycin, the aminoglycoside G418,

hygromycin, or the like. The particular marker employed is not essential to this invention, with a particular marker preferred depending on the particular host and the manner of construction.

For *Agrobacterium*-mediated transformation of plant cells, explants are incubated with *Agrobacterium* for a time sufficient to result in infection , the bacteria killed, and the plant cells cultured in an appropriate selection medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of the recombinant protein produced by the plants.

There are a number of possible ways to obtain plant cells containing more than one expression construct. In one approach, plant cells are co-transformed with a first and second construct by inclusion of both expression constructs in a single transformation vector or by using separate vectors, one of which expresses desired genes. The second construct can be introduced into a plant that has already been transformed with the first expression construct, or alternatively, transformed plants, one having the first construct and one having the second construct, can be crossed to bring the constructs together in the same plant.

A. Plants

Host cells of the present invention include plant cells, both monocotyledenous and dicotyledenous. In one embodiment, the plants used in the methods of the present invention are derived from monocots, particularly the members of the taxonomic family known as the *Gramineae*. This includes all members of the grass family of which the edible varieties are known as cereals. The cereals include a wide variety of species such as wheat (*Triticum sps.*), rice (*Oryza sps.*) barley (*Hordeum sps.*) oats, (*Avena sps.*) rye (*Secale sps.*), corn (maize) (*Zea sps.*) and millet (*Pennisettum sps.*). In practicing the present invention, preferred grains are rice, wheat, maize, barley, rye, triticale. Also preferred are dicots exemplified by soybean (*Glycine* spp.). However, the heterologous expression constructs herein can be used in any monocots or dicots, whether edible or not, for production of any heterologous polypeptides whether for human or animal consumption or use, whether for health benefits or for industrial purposes.

Plant cells or tissues are transformed with an expression vector comprising the coding sequence for a desired heterologous polypeptide. Transgenic plant cells obtained

as a result of such transformation express the coding sequence for the heterologous polypeptide, exemplified herein as lysozyme, lactoferrin or α_1-antitrypsin. The transgenic plant cells are cultured in medium containing the appropriate selection agent to identify and select for plant cells which express the heterologous nucleic acid sequence. After plant cells that express the heterologous nucleic acid sequence are selected, whole plants are regenerated from the selected transgenic plant cells. Techniques for regenerating whole plants from transformed plant cells are generally known in the art. Transgenic plant lines, *e.g.,* rice, wheat, corn or barely, can be developed and genetic crosses carried out using conventional plant breeding techniques.

Production of recombinant proteins in seeds, *e.g.*, rice (*Oryza sativa* L.) seeds, has the advantages that (a) high level expression makes it an economically practical strategy, and (b) rice is a normal part of the diet of infants and children, has good nutritional value and low allergenicity. Thus, the use of rice as the basis for a food supplement containing the heterologous polypeptide is unlikely to introduce any risk associated with animal pathogens and thereby eliminates the need for a high degree of purification, for example, when included in infant formula.

In addition, rice is the staple food crop of more than half the world's population. Recent reports on the production of provitamin A (beta-Carotene) in rice seeds exemplifies the unmet need for value-added food crops, especially in the developing world (Ye *et al.*, 2000) where rice is used as major food crop.

VI. Detecting Expression of Recombinant Heterologous Proteins

Transformed plant cells are screened for the ability to be cultured in selective media having a threshold concentration of a selective agent. Plant cells that grow on or in the selective media are typically transferred to a fresh supply of the same media and cultured again. The explants are then cultured under regeneration conditions to produce regenerated plant shoots. After shoots form, the shoots are transferred to a selective rooting medium to provide a complete plantlet. The plantlet may then be grown to provide seed, cuttings, or the like for propagating the transformed plants. The method provides for efficient transformation of plant cells with expression of a gene of autologous or heterologous origin and regeneration of transgenic plants, which can produce a recombinant human milk protein.

The expression of the recombinant heterologous polypeptide may be confirmed using standard analytical techniques such as Western blot, ELISA, PCR, HPLC, NMR,

or mass spectroscopy, together with assays for a biological activity specific to the particular protein being expressed.

Example 3 describes the characterization of human lysozyme produced in the seeds of transgenic rice plants. Analyses used to confirm that recombinant lysozyme produced in transgenic rice is essentially the same as the native form of the protein both in physical characteristics and biological activity included, SDS-PAGE, reverse IEF gel electrophoresis, Western blot analysis, enzyme linked immunosorbant assay (ELISA), enzymatic activity assay and bactericidal activity assay using indicator strains, *Micrococcus luteus* and *E. coli* strain JM109.

Example 4 describes the characterization of human lactoferrin produced in the seeds of transgenic rice plants. Analyses used to confirm that recombinant lactoferrin produced in transgenic rice is essentially the same as the native form of the protein both in physical characteristics and biological activity included, Southern blot, Western blot, ELISA, N-Terminal Amino Acid Sequencing, analysis of glycosylation and determination of sugar content, a determination of the isoelectric point, pH dependent iron release of rLF, bacteriostatic activity assay of rLF using enteropathogenic *E. coli* as the indicator strain.

Example 5 details the characterization of α_1-antitrypsin produced by transgenic monocot plant cells, and Example 6 shows other milk proteins also produced by monocot plant transformed with the chimeric genes of the invention.

VII. Preparation of Seed Compositions, and Methods of their Use

The invention provides, in one aspect, a protein-containing product composed of a monocot seed fluor, extract, or malt composition containing monocot seed or malted-seed components, one or more seed-produced, non-seed proteins in substantially unpurified form, and a vehicle containing the extract in a form suitable for human or animal use. For use in a food or feed product, the vehicle may be a capsule, binder components effective to tabletize the composition, a consumable liquid, or a consumable suspension. The vehicle may be a processed food in which the extract is mixed. Below are described methods for preparing the flour, extract, or malt compositions.

A. Flour composition

The flour composition is prepared by milling mature monocot plant seeds, using standard milling and, optionally, flour purification methods, e.g., in preparing refined

flour. Briefly, mature seeds are dehusked, and the dehusked seeds then ground into a fine flour by conventional milling equipment.

The flour may be added to foods during food processing according to standard food processing methods. Preferably, the processing temperature does not lead to denaturation of the milk proteins, *e.g.*, above 60°-70°C. The flour may also be used directly, either in capsule, tabletized, or powder form, as a neutriceutical composition. For producing cosmetic or care products, such as toothpaste or topical creams, the flour may be blended with other components of the product (the vehicle). For preparing a surgical dressing or surgical powder, e.g., for blot clotting, where the heterologous protein is fibrinogen or other clotting factor(s), the vehicle is a surgical dressing or container for delivering the powder.

One preferred flour composition contains lactoferrin and/or lysozyme. The flour may alternatively, or in addition, include one or more of the other human milk proteins such as epidermal growth factor, insulin-like growth factor-1, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, and α_1-antitrypsin. Alternatively, the flour may contain another type of therapeutic protein, as given above, a clotting factor, an industrial enzyme, or other protein useful in a substantially unpurified state in a monocot-seed matrix.

Flour containing two or more heterologous polypeptides may be prepared by combining flour from seeds that separately produce the different proteins, for example, equal amounts of a flour containing lysozyme and a flour containing lactoferrin. Alternatively, a multi-protein composition can be prepared as seed flour from plants, such as monocot plants co-transformed with chimeric genes expressing different milk proteins, e.g., lactoferrin and lysozyme.

B. An extract

An extract composition is prepared by milling seeds to form a flour, extracting the flour with an aqueous buffered solution, and optionally, further treating the extract to partially concentrate the extract and/or remove unwanted components. Details of exemplary methods for producing the extract composition are given in Example 11. Briefly, mature monocot seeds, such as rice seeds, are milled to a flour, and the flour then suspended in saline or in a buffer, such as Phosphate Buffered Saline ("PBS"), ammonium bicarbonate buffer, ammonium acetate buffer, Tris buffer or a volatile buffer that would evaporate upon drying. The flour suspension is incubated with shaking for a period typically between 30 minutes and 4 hours, at a temperature between 20-55°C.

The resulting homogenate is clarified either by filtration or centrifugation. The clarified filtrate or supernatant may be further processed, for example by ultrafiltration or dialysis or both to remove contaminants such as lipids, sugars and salt. Finally, the material may dried, e.g., by lyophilization, to form a dry cake or powder. The extract has the advantage of high recombinant polypeptide yields, limiting losses associated with protein purification. At the same time, the recombinant heterologous polypeptides are in a form readily usable and available upon ingestion of the extract or food containing the extract.

One particular advantage of the extract is the low amount of seed starch present in the extract. In particular, the extract may increase the concentration of recombinant protein, from about 0.5% of total soluble protein ("TSP") in conventional approaches to over about 25% of TSP in the extract approach. Some of the present extract approach even reached 40% of TSP depending on the expression level of the recombinant protein in the seeds. In addition, the extract approach removes starch granules, which require high gelling temperature, for example above about 75°C. Consequently, the extract approach provides more flexibility in processing the transgenic rice or other transgenic grain containing the recombinant polypeptides into food and nutritional drinks.

The extract can be used as a nutraceutical for direct use, e.g., in capsule, tabletized or powder form, or as a food additive in food processing. In one embodiment, the extract is added to an infant milk formula, in an amount typically between 0.1 to 10 percent by dry weight, preferably 1-5% by dry weight of the total formula weight. One preferred infant formula contains both lactoferrin and lysozyme, preferably in an amount between 50-200% of the amount of human lactoferrin or lysozyme, respectively, of that found in normal human milk. As noted above, lactoferrin is present in a concentration of about 1 gram /liter human milk, and lysozyme, about 0.1/liter human milk.

The extract may also be used directly, either in capsule, tabletized, or powder form, as a neutriceutical composition, where the product vehicle is the capsule, container, or binder for the composition. For producing cosmetic or care products, such as toothpaste or topical creams, the extract may be blended with other compoents of the product (the vehicel). For preparing a surgical dressing or surgical powder, e.g., for blot clotting, where the heterologous protein is fibrinogen or other clotting factor(s), the vehicle is a surgical dressing or container for delivering the powder.

One preferred flour composition contains lactoferrin and/or lysozyme. The flour may alternatively, or in addition, include one or more of the other human milk proteins such as epidermal growth factor, insulin-like growth factor-1, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, and α_1-antitrypsin. Alternatively, the flour may contain

another type of therapeutic protein, as given above, a clotting factor, an industrial enzyme, or other protein useful in a substantially unpurified state in a monocot-seed matrix.

As above, extract containing two or more heterologous polypeptides may be prepared by combining extracts from seeds that separately produce the different proteins, or by processing seeds from plants co-transformed with chimeric genes expressing different milk proteins, e.g., lactoferrin and lysozyme.

C. Malt composition

In accordance with another embodiment, the invention provides a malt extract or malt syrup ("malt") composition in which seed starches have been largely reduced to malt sugars, and the milk protein(s) are in an active, bioavailable form. A wide range of food and feed products, including processed foods, food supplement, sweetener, and feed additive may be produced by varying the types of malt used, the mashing program and the ways in which the wort is subsequently handled. If materials other than barley malt are used in the mash (such as starch from other grains). The resulting product is classified as a malt syrup.

Malt extracts, which may have a syrupy consistency or may be powders, are made by mashing ground malt, usually barley malt, in conventional brewery equipment, collecting the wort and concentrating it or drying it. Modern production of food malt extracts and malt syrups has evolved into three basic grain stages: steeping, germination, and drying of the germinated seed, followed by three more steps involving liquefaction of the germinated grain, mashing of the germinated grain, lautering (filtering), and evaporation. Many variations of malt extracts or syrups are possible. Flavor, color, solids, enzymatic activity, and protein are the basic characteristics that can be adjusted during production to provide malts specific for given food applications. (See, generally, Eley; Hickenbottom, 1996, 1997a, 1997b, 1983; Lake; Moore;Moe; Sfat; Doncheck; Briggs, 1981, 1998; and Hough).

(i) Steeping

After the barley of choice has been cleaned of foreign material, it is graded to size and transferred to steep tanks equipped with water inlet and outlet pipes. Compressed air is fed from the tank bottom for vigorous aeration and mixing for the barley/water mixture. When the barley has reached a water content of 43-45%, steeping is stopped.

(ii) Germination

The steeped barley is moved to germination floors or rooms depending on the particular malt house's capabilities and allowed to germinate under controlled temperature, air, and moisture conditions. Total germination varies from four to seven days, depending on the barley type, density end use of the malt, and the controls or germination method used. All aspects of germination must be kept in constant balance to ensure proper kernel modification and yield.

Many enzymatic systems are activated during germination. Two of the systems are the oxidative and reductive systems involved with the respiration phase. Other enzymes break down the endosperm cell structure, which in itself if a measure of germination rate when the pentose production is evaluated. The proteolytic enzymes release or active beta-amylase and also work on the proteins present to render them soluble. In fact, about 40% of the total protein is made soluble in water. Optimum germination activates a balanced enzyme system, which hydrolyzes the starch present.

(iii) Kilning

Drying or kilning, when done at the proper time and optimum degree of starch modification, stops the germination. The heat also catalyzes additional reactions, notably flavor and color development. When drying is complete, the sprouts and other extraneous materials are removed, and the kernels are then ready for further processing.

(iv) Malt Extracts And Syrups

The malted barley (kernel) is coarsely ground in crushers and fed into mash tuns where it is mixed with water. During a series of time and temperature changes, some of the starch is converted into fermentable sugars by action of the natural alpha- and beta-amylases, better known as the diastatic system. If cereal adjuncts are to be added, which result in malt syrups with mellower and sweeter flavors than the extracts, they are added at this stage usually derived from the cereal grains, corn and rice, although barley, wheat, rye, millet and sorghum are sometimes used, derived from mature seeds that produce the desired recombinant milk proteins.

Once the mash batch has achieved the correct degree of hydrolysis, it is transferred to lauter tuns. The lauter tun has a slotted or false bottom a few inches above the real bottom to allow for filtration and is also equipped with some means of agitation. During this extraction stage, the amyolytic enzymes liquefy additional insoluble starches, converting them to maltose and dextrins. At the same time, the proteolytic enzymes attach certain proteins converting them into simpler, soluble forms. After the appropriate

conditions have been met, the liquid phase, or wort, is drawn from the lauter tuns into evaporators.

Evaporation of the wort is conducted under vacuum where it is converted into a syrup of about 80% solids. Depending on the temperatures used, malt extracts or syrups of high, medium, or zero enzymatic activity can be produced. Color and flavor also can be controlled during this stage. The finishing steps of filtering, cooling, and packaging complete the malt extract/syrup process.

(v) Transgenic Malt Extract

For a transgenic malt extract, the starting barley is a transgenic barley engineered to produce on or more human milk proteins in the endosperm either in grain maturation or in the malting process, or at both times. Malting and processing times and conditions are adjusted so that the bioactivity of the target recombinant molecules is preserved and the bioavailability of the target recombinant molecule is maximized. The resulting malt extract is either consumed directly as a concentrate that is either consumed directly as a food, or is incorporated as an ingredient in a food mixture. Studies conducted in support of the present invention demonstrate that recombinant proteins retain activity after malting for up to at least 288 hrs.

(vi) Transgenic Malt Syrup

For a transgenic malt syrup, the starting barley can be a non-transgenic barley, or a transgenic barley, or a mixture of both. The barley is processed as described, except that during the mashing process, a cereal adjunct is added in a form that it is converted during the mashing process with the concurrent retention and generation of bioavailability and bioactivity of the target recombinant molecule fond within the transgenic cereal adjunct. The use of a transgenic cereal adjunct enables the production in the malt syrup of the target recombinant molecule expressed in the transgenic grain endosperm.

The malt extract or syrup may be used directly as a syrup, or added to processed foods or drinks, according to standard food processing procedures that employ grain extracts or syrups, e.g., for sweetening. One food is an infant formula containing between 0.1 to 10% malt (extract or syrup). The malt is also useful as a sweetener/nutritional additive in baby and adult foods, and nutritional drinks.

Preferred malt extracts or syrups contain lactoferrin and/or lysozyme. The malt may alternatively, or in addition, include one or more of the human milk proteins such as epidermal growth factor, insulin-like growth factor-1, lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin and immunoglobulins. Malts containing

other proteins, e.g., therapeutic proteins, which may be delivered orally are also contemplated. As above, malt containing two or more milk proteins may be prepared by combining or preparing malts from seeds that separately produce the different proteins, or by preparing a malt from the seeds of plants co-transformed with chimeric genes expressing different milk proteins, e.g., lactoferrin and lysozyme.

D. Seed-Composition Products and Methods

A variety of products containing a flour, extract, or malt seed composition in a suitable vehicle are contemplated, as given below:

(i) Nutritional supplement, sweetener, food, or feed. In this general embodiment, the seed composition contains an orally active nutritional peptide, such as a milk protein, e.g., lysozyme and/or lactoferrin, in a suitable fluor, extract, or malt composition, or an anti-microbial protein. The composition is carried in a suitable vehicle which may be a capsule or binder or liquid suspension, for direct ingestion, a container where the composition is used as a sweetener or other additive, or a food or feed, where the composition is used to produce a nutritionally enhanced food or feed.

(ii) Surgical dressing, bandages, and clotting compositions. In this embodiment, the seed composition contains a blood-clotting factor, such as fibrinogen in a suitable fluor, extract, or dried malt composition. For a bandage product, the vehicle is the bandage itself, carrying the composition in a dried form, e.g., in a bandage matrix. Where the composition is used directly as a clotting agent for surgical clotting, the vehicle is a suitable dispenser for delivering the composition in dried or liquid form to a surgical wound.

(iii) Industrial enzyme or detergent product. The composition may contain an industrial enzyme, such as a protease, useful in detergent or other cleaning or processing products. Here the composition is preferably an extract or fluor added directly to the product of interest.

(iv) Anti-infective agent. The composition may contain a seed-produced antimicrobial agent, such as the anti-microbial milk proteins as lactoferrin, lysozyme, lactoferricin, lactohedrin, kappa-casein, haptocorrin, or lactoperoxidase, a milk protein like alpha-1-antitrypsin that may function as an antimicrobial by inhibiting proteolysis of other anti-microbial proteins, non-milk anti-microbial proteins and acute-phase proteins, e.g., proteins that are produced normally in production animals in response to infection, such as C-reactive protein, serum amyloid A; ferritin, haptoglobin, seromucoids, ceruloplasmin, 15-keto-13,14-dihydro-prostaglandin F2 alpha, fibrinogen, alpha-1-acid

glycoprotein, mannose binding protein, lipopolysaccharide binding protein, alpha-2 macroglobulin and defensins, anti-microbial protein(s), such as cecropin, magainin, defensins, tachyplesin, parasin I, buforin I, PMAP-23, moronecidin, anoplin, gambicin, and SAMP-29, or other anti-microbial proteins, such as .CAP37, granulysin, secretory leukocyte protease inhibitor, CAP18, ubiquicidin, bovine antimicrobial protein-1, Ace-AMP1, tachyplesin, big defensin, Ac-AMP2, Ah-AMP1, and CAP18..

The composition may be added to food or feed, to confer enhanced resistance to microorganisms, or may be used directly as a nutraceutical product, e.g., in capsule form, for oral administration. The composition containing the anti-microbial protein may be incorporated, typically in fluor or extract form, into cosmetics, bandages, or other health care products to confer anti-microbial properties to the product.

From the foregoing, it can be appreciated how various objects and features of the invention are met. The production of high levels of heterologous polypeptide in grains, exemplified herein by rice, provides the distinct advantage that protein-containing compositions useful in a variety of products may be prepared with little or no purification.

Transgenic seeds are ideal bioreactors, combining low production costs and low or minimal downstream processing costs prior to use. Seed grain proteins can accumulate to 9-19% of grain weight (Lásztitym 1996); the endosperm proteins are synthesized during grain maturation and stored in protein bodies for use in the germination and seedling growth of the next plant generation; grains can be stored for years without loss of functionality, and therefore the downstream processing can be conducted independently of growing seasons.

The results presented herein demonstrate that human milk proteins may be expressed at high levels in the seeds of transgenic plants, e.g., up to 0.25 to 1% of total seed dry weight. The production of high levels of human milk proteins in grains, exemplified herein by rice, provides the distinct advantage that food supplements may be prepared with little or no purification. Since the recombinant grain finds utility as a food or food supplement, as a flour, extract or malt, the regulatory requirements for purity are not stringent.

All publications, patents and patent applications are herein expressly incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

The following examples illustrate but are not intended in any way to limit the invention.

EXAMPLE 1

Expression Vectors for Generation of Transgenic Plants

In general, expression vectors were constructed using standard molecular biological techniques as described in Ausubel *et al.*, 1987. The vectors contain a heterologous protein coding sequence for lactoferrin or lysozyme under the control of a rice tissue-specific promoter, as further described below.

A. An Expression Vector For Human Lysozyme Expression In Transgenic Rice Cells

The synthesized lysozyme gene was cloned into an API base vector pAPI137 by conventional molecular cloning techniques (Sambrook *et al.*, 1989). Plasmid pAPI137 contains the *RAmy*3D promoter (Huang *et al.*, 1993), the codons for the *RAmy*3D signal peptide and the *RAmy*3D terminator. The *RAmy*3D promoter, isolated from the rice amylase gene family, is activated in rice calli by sugar starvation (Huang *et al.*, 1993). The human lysozyme gene was placed between the sequences of the *RAmy*3D signal peptide and the *RAmy*3D terminator to give plasmid pAPI156 having a size of 4829 bp.

The promoter of the rice Glutelin 1 gene (*Gt-1*) and the nucleotide sequence of the signal peptide were cloned with two primers based on the published Gt1 gene sequence (Okita et al. J Biol Chem 264: 12573-12581, 1989). The forward primer with *Hind*III site was named MV-*Gt-1*-F1; 5'-ATCGAAGCTTCATGAGTAATGTGTGAGCATTATGGGACCACG-3' (SEQ ID NO:5). The reverse primer was named Xba-*Gt-1*-R1; 5'-CTAGTCTAGACTCGAGCCACGGCCATGGGGCCGGCTAGGGAGCCATCGCACAAGAGGAA-3' (SEQ ID NO:6). Genomic DNA was isolated from leaves of rice variety M202 (Dellaporta et al., 1983). The PCR product amplified from the genomic DNA was cloned into pCR 2.1 (Invitrogen, Carlsbad, CA). The resulting plasmid was named pCR*Gt-1* or pAPI134.

To generate a *Gt-1* expression plasmid, pAPI134 was digested with *Hind*III and *Xba*I. The fragment containing the *Gt-1* promoter and *Gt-1* signal peptide was cloned into a pUC19 based plasmid containing the nopaline synthase 3' (*nos*) terminator. The resulting plasmid was named pAPI141 and contains the rice *Gt-1* promoter, the *Gt-1* signal peptide, a multiple cloning site and the *nos* terminator.

The synthesized human lysozyme gene "lys-ger" (by Operon Technologies, Inc., Alameda, CA) that was optimized based on the rice gene codon usage was digested with *Dra*I and *Xho*I and cloned into pAPI141 digested with *Nae*I and *Xho*I according to standard cloning techniques (Sambrook et al., 1989). The resulting plasmid was called pAPI159 (Fig. 1) having a size of 4131bp.

B. An Expression Vector For Human Lactoferrin Expression In Transgenic Rice

The hLF gene (Rey, MW, 1990) was codon optimized and synthesized by Operon Technologies (CA, USA). The plasmid containing the codon-optimized gene was called Lac-ger. Lac-ger was digested with *Sma*I/*Xho*I and the fragment containing the lactoferrin gene was cloned into pAPI141 which was partially digested with *Nae*I and completely digested with *Xho*I. The resulting plasmid was named pAPI 164. For expression of hLF in rice seeds, the codon optimized gene was operably linked to the rice endosperm specific glutelin (Gt1) promoter and NOS terminator (Fig. 7).

EXAMPLE 2

Generation Of Transgenic Plant Cells Expressing Human Milk Proteins

The procedure of microprojectile-mediated rice transformation (US Patent 6,284,956) was followed. Calli was raised from TP309 mature rice seeds, with calli two to four mm in diameter selected and placed on N6 media supplemented with 0.3 M mannitol and 0.3 M sorbitol for 20 hours before bombardment. Biolistic bombardment was carried out with the biolistic PDC-1000/He system (Bio-Rad, USA). Plasmid carrying milk protein genes and pAPI76, a plasmid carrying hygromycin selectable marker gene were gold-coated and co-bombarded at a ratio of 6:1 with a helium pressure of 1100psi. Two day old bombarded calli were then transferred to N6 selection media supplemented with 20 mg/l hygromycin B and allowed to grow in the dark at 26°C for 45 days.

In order to develop transgenic rice plants, the selected calli were transferred to pre-regeneration and regeneration media. When regenerated plants became 1-3 cm in height, the plantlets were transferred to rooting media which consisted of half concentration of MS and 0.05 mg/l NAA. After two weeks, plantlets with developed roots and shoots were transferred to soil and kept under the cover of plastic container for a week. The plants were allowed to grow about 12 cm tall and shifted to the green house where they were grown up to maturity.

A. A Plasmid for Hygromycin Selection

Plasmid, pAPI76, carrying the bacterial hygromycin B phosphotransferase (hpt) gene was made as described in U.S. Patent No. 6,284,956, and was used for co-transformation of calli to allow selection of the transformants. Briefly, pAPI176 was created as follows: A DNA fragment was amplified from a rice α-amylase gene, *RAmy*1A (Huang *et al.*, 1990b) and cloned into pBluescript KS+ at the *Sma*I/*Eco*RI restriction sites and the resulting plasmid was called p1AT. The PCR amplified fragment contained 297 bp of *RAmy*1A terminator. A *Bam*HI DNA fragment from pGL2 (Shimamoto *et al.*, 1989) was cloned into the *Bam*HI site of p1AT and the resulting plasmid was called pAPI174. Finally a *Sac*I/*Xba*I fragment amplified from glucanase gene, *Gns*9 (Romero *et al.*, 1998) was inserted into pAPI174 using the same restriction sites. The PCR generated, *Gns*9 promoter fragment was confirmed by DNA sequencing. The resulting plasmid was named as pAPI176.

B. Generation of Human Lysozyme Expressing Transgenic Rice Cells and Plants

The synthetic human lysozyme (hLys) gene under the control of the *RAmy*3D promoter and terminator in the pAPI156 plasmid (example 1A) was used to generate sixty independent transformants by particle bombardment-mediated transformation.

Particle bombardment mediated transformation of rice was carried out as described above. Briefly, rice calli derived from TP309 were bombarded with gold particles coated with plasmids pAPI156 and pAPI76 in a ratio of 6:1 using the helium biolistic particle delivery system, PDS 1000 (Bio-Rad, CA). Transformed calli were selected in the presence of hygromycin B (35 mg/L) on N6 (Sigma, MO).

Selected cell lines were maintained in culture media with 3% sucrose (Huang *et al.*, 1993). Lysozyme expression was induced by sugar starvation. Briefly, AA medium (containing 3% sucrose) was removed by aspiration, followed by washing the cells three times with AA minus sucrose (AA-S). The cells were then incubated with AA-S at 40% (v/v) density for three and a half days to obtain the optimal level of lysozyme expression.

Transformants expressing lysozyme were identified by immunoblot analysis, turbidimetric rate determination with *Micrococcus lysodeikticus* or ELISA. Calli were ranked according to the expressed lysozyme level. Suspension cell cultures from the top lines were established following the procedure described previously (Huang *et al.*, 1993). The amount of total protein (Bradford assay) and lysozyme (ELISA) was evaluated in selected calli (Table 1).

Table 1. Expression Level Of Human Milk Lysozyme In Transformed Calli

Cell line	Calli (g)	Total protein (µg)	Lysozyme (µg)	Lysozyme/protein (%)
156-1	0.39	2626.5	65.7	2.5
156-5	0.38	5510	68.9	1.25
156-16	0.4	4815	120.4	2.5
156-19	0.44	2440	30.5	1.25
156-28	0.49	4910	24.6	0.5
156-43	0.56	8150	101.9	1.25
156-47	0.37	2472	6.2	0.25

The synthetic human lysozyme (hLys) gene under the control of the *Gt1* promoter and Nos terminator in the pAPI159 plasmid (Fig. 1) was used to generate independent transformants by particle bombardment-mediated transformation. Transformed calli were selected as described above, then transferred to pre-regeneration and regeneration media. When regenerated plants became 1-3 cm in height, the plantlets were transferred to rooting media which consisted of half concentration of MS and 0.05 mg/l NAA. After two weeks, plantlets with developed roots and shoots were transferred to soil and kept under the cover of plastic container for a week. The plants were allowed to grow about 12 cm tall and shifted to the green house where they were grown up to maturity (R0 plants).

Screening for R0 plants expressing human lysozyme. Individual rice endosperms or grains were ground with cold phosphate buffered-saline (PBS) with the addition of 0.35 M NaCl. Grinding was conducted with a pre-cooled mortar and pestle at 1 ml buffer/grain. Clear grain homogenate was obtained by subjecting the resulting grain extract to centrifugation at 14,000 rpm for 10 min at 40°C.

Embryos from individual R_1 seed (derived from R_0 plants) that showed a level of lysozyme expression that was greater than 10 µg/seed were saved and used to generate R1 plants. Briefly, seeds were dissected into embryo and endosperm portions. The endosperm was ground and assayed for lysozyme expression (as further described below). Embryos were sterilized in 50% commercial bleach for 25 minutes and washed with sterile H_2O three times for 5 minutes each. Sterilized embryos were placed in a tissue culture tube that contained MS solid medium. Embryos germinated and plantlets having about three inches shoots and healthy root systems were obtained in two weeks. The plantlets were then transferred to pots to obtain mature plants (R_1).

A total of 197 embryos from 12 selected R0 plants were germinated and 157 R_1 seedlings planted in the greenhouse for generation of R2 grains. Individual R2 grains (n=1502) from 109 R1 fertile plants were screened for lysozyme expression by lysozyme activity assay in order to identify 42 homozygous plants.

Homozygous R1 plants were identified by analyzing positive expressions of recombinant human lysozyme (rHlys) from a minimum of 20 individual R2 grains. Homozygous lines derived from these plants were planted in a rice field in California. During growth, agronomic characteristics of both transgenic and non-transgenic plants, such as plant height, percentage of fertility, number of effective tillers, filled grains/panicle, non-filled grains/plant, time to maturity and 1000 grain weight were determined and compared. Plants with satisfactory agronomic traits were selected and rHlys expression levels were determined by lysozyme activity assay. Plants that met the criteria for satisfactory agronomic traits and had more than 35 µg of rHlys/grain were advanced to next generations.

SDS-PAGE, electroblotting and Western blot analysis were carried out with 18% precast gel (Invitrogen, Carlsbad, CA) as described in Example 3. The primary rabbit polyclonal antibody against human lysozyme was purchased from Dako A/S (Denmark) and used at 1:5000. Lysozyme was quantified by a turbidimetric activity assay with *Micrococcus luteus* (Sigma) on 96-well microtiter plate as described in Example 3. Briefly, 250 µl of 0.015% *M. luteus* cell suspension was incubated with 10 µl of samples containing lysozyme with a concentration less than 2.4 µg/ml. The reaction was followed by the kinetic mode in Microplate Manager (Bio-Rad, CA) for 5 min at 450 nm. The concentration of lysozyme was then determined in reference to the standard curve.

The stable expression level of human lysozyme (rHlys) reached at least about 0.6% rHlys per brown rice weight amounting to 45% of the total soluble protein extract from rice grain. Figure 2 illustrates the seed specific expression of human lysozyme in transgenic plants. rHlys is only found in mature and germinated grain, but not in any other tissues tested. Fig. 6 shows the expression level of human lysozyme in powdered R3 seeds taken from transgenic rice plants

C. Generation of Human Lysozyme-Expressing Transgenic Wheat Cells and Plants

Plasmid API159 (Figure 1) and API230 (Figure 35) were used to transform wheat cells substantially in the same manner as in transforming rice cells. Eight transgenic wheat lines were produced with API159, generating an expression level of about 150 to

300 μg of lysozyme per grain. Two transgenic wheat lines were produced with API230, yielding an expression level of about 50 to 120 μg of lysozyme per grain.

D. Generation of Human Lysozyme-Expressing Transgenic Barley Cells and Plants

The plasmid API159 was also used to transform barley cells substantially as described as transformation of rice cells. Five transgenic barley lines were produced, yielding about 3.9 to 12.3 μg of lysozyme per grain.

E. Generation of Human Lactoferrin Expressing Transgenic Rice Cells and Plants

The synthetic human lactoferrin gene under the control of the Gt1 promoter in the pAPI164 plasmid was used to generate over 100 independent transformants by particle bombardment-mediated transformation.

Particle bombardment mediated transformation of rice was carried out as described above. At least 20 R1 grains from each R0 plant were analyzed for rHLF expression. Individual R1 grains were cut into halves. The endospermic half was subjected to rHLF expression analysis by Western blot or ELISA and the corresponding positive embryonic half was germinated to generate R1 seedlings. The seedlings were transplanted to generate R2 grains. During the screening of R1 grains we observed that all the positive grains were opaque-pinkish in color in comparison to negative or control grains. The opaque-pinkish color in rice grains was then used to identify homozygous lines. A transgenic plant was considered to be homozygous and expressing rHLF if all grains from that plant were opaque-pinkish. Homozygous lines were then confirmed by ELISA analysis. Based on the expression analysis and agronomic characters, selected homozygous R2 lines were advanced to R3 and R6 generations.

EXAMPLE 3

Characterization of Recombinant Human Lysozyme (rLys) Produced by Transgenic Rice Cells and Plants

A. Southern blot analysis

About three grams of young leaves were collected and grounded with liquid nitrogen into a fine powder. The genomic DNA was isolated according to the procedure as described in Dellaporta *et al.*, 1983, and purified by phenol-chloroform extraction. Approximately 5 μg of DNA was then with *Hind*III and *Eco*RI, separated on a 1%

agarose gel, blotted onto a Hybond$^+$ membrane (Amersham Pharmacia Biotech, Piscataway, NJ). The blot was probed with gel purified human Hlys gene and developed by ECL^{TM} direct nucleic acid labeling and detection system (Amersham Pharmacia). By comparing to known amounts of the intact 1470 bp human lysozyme (Hlys) gene, the intact copy number of the transgenes, including promoter and Hlys gene, was estimated to vary from about 1 to about 6. No positive correlation between copy number of the rHlys transgene and amount of rHlys synthesized was discernible.

B. SDS-PAGE and reverse IEF gel electrophoresis

Induced calli or harvested cells from suspension cell cultures were ground with cold phosphate buffered-saline (PBS) with a protease inhibitor cocktail (2 µg/ml aprotonin, 0.5 µg/ml leupeptin, 1 mM EDTA and 2 mM Pefabloc). The protease inhibitor cocktail was excluded from the buffer used subsequently during the purification of the enzyme, since the inhibitors did not increase the lysozyme expression yield. Grinding was conducted with a pre-chilled mortar and pestle at approximately 2 ml buffer/g calli or cells. A clear homogenate was obtained by subjecting the resulting extract to centrifugation at 16,000 x g for 10 minutes at 4 C.

SDS-PAGE was carried out using an 18% precast gel (Novex, CA). The resulting gel was stained with 0.1% Coomassie brilliant blue R-250 at 45% methanol and 10% glacial acetic acid for three hours. Gel destaining was conducted with 45% methanol and 10% glacial acetic acid until the desired background was reached.

Reverse IEF gel electrophoresis was carried out using a precast Novex pH 3-10 IEF gel according to the manufacturer's instructions (Novex, CA). About 30 µg of lysozyme was loaded onto the gel and electrophoresed at 100 V for 50 minutes followed by application of 200 V for 20 minutes. The gel was then fixed in 136 mM sulphosalicylic acid and 11.5% TCA for 30 minutes and stained in 0.1% Coomassie brilliant blue R-250, 40% ethanol, 10% glacial acetic acid for 30 minutes. The destaining solution contained 25% ethanol and 8% acetic acid.

C. Western blot analysis

A SDS-PAGE gel was electroblotted to a 0.45 µm nitrocellulose membrane using a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad, CA) and subsequently subjected to immuno-blotting analysis. The blot was blocked with 5% non-fat dry milk in PBS, pH 7.4 for at least two hours followed by three washes with PBS, pH 7.4 for 10 minutes each. The primary rabbit polyclonal antibody against human lysozyme (Dako

A/S, Denmark) was diluted at 1:2000 in the blocking buffer and the blot was incubated in the solution for at least one hour. The blot was then washed with PBS three times for 10 minutes each. The secondary goat anti-rabbit IgG (H+L)-alkaline phosphatase conjugate (Bio-Rad, CA) was diluted in the blocking buffer at 1:4000. The membrane was then incubated in the secondary antibody solution for one hour and then washed three times. Color development was initiated by adding the substrate system BCIP-NBT (Sigma) and the process was stopped by rinsing the blot with H_2O once the desirable intensity of the bands had been achieved.

D. Enzyme Linked Immunosorbant Assay (ELISA)

An indirect sandwich ELISA was developed to quantify total lysozyme expressed in rice calli or cells and used as an alternative assay to determine the lysozyme expression yield. A direct sandwich ELISA for lysozyme quantification has been previously reported (Lollike *et al.*, 1995, Taylor, 1992), however an alternate assay was developed as a key reagent used in the assay is no longer commercially available.

In carrying out the assay, rabbit anti-human lysozyme antibody (Dako D/K, Denmark) was used to coat a 96 well plate at 1:5000 diluted in PBS overnight at room temperature. After washing with PBS, the plate was blocked with 5% normal donkey serum (Jackson ImmunoResearch Laboratories, PA) in PBS for one hour. The plate was washed again with PBS. Lysozyme samples were diluted in 0.05% Tween in PBS and captured by adding to the plate and incubating for one hour. After washing the plate with PBS, sheep anti-human lysozyme at 1:1000 diluted with 0.05% Tween in PBS was added and incubated for one hour. The plate was washed again with PBS. Peroxidase-conjugated affinipure donkey anti-sheep IgG (H+L) diluted in 0.05% Tween in PBS at 1:10,000 was added and incubated for one hour. After a final wash of the plate with PBS, color was developed by incubating the plate with TMB substrate (Sigma, MO) for 5-15 minutes and the absorbance read at 655 nm.

E. Enzymatic Activity Assay For Lysozyme

A reliable and quantitative method was developed to analyze the expression level of enzymatically active lysozyme. The turbidimetric assay was developed using a 96-well microtiter plate format and based on the standard lysozyme assay that is carried out spectrophotometrically in cuvettes. A microtiter plate based method previously described for the detection of lysozyme release from human neutrophils had a detection

range of 1-100 ng/ml (Moreira-Ludewig *et al.*, 1992). The assay conditions were modified to maintain the linearity of detection up to 3.0 µg/ml.

The enzymatic activity of lysozyme was routinely determined by spectrophotometric monitoring of the decrease in turbidity at 450 nm of a suspension of *Micrococcus luteus (M. lysodeikticus)* cells (Shugar, 1952). Specifically, 250 µl of a 0.015% (w/v) *Micrococcus leteus* cell suspension was prepared in 66 mM potassium phosphate, pH 6.24 (buffer A). Cell suspensions were equilibrated at room temperature and the reaction was initiated by adding 10 µl samples containing lysozyme with concentrations from 0 to 2.4 µg/ml. Lysozyme activity was determined in a kinetic mode for 5 minutes at 450 nm. The concentration of lysozyme was then calculated by reference to the standard curve constructed with human milk-derived lysozyme.

The enzymatic activity of human milk lysozyme and the rice cell derived lysozyme of the invention was compared. As shown in Figure 4, the lysozyme effected reduction of the turbidity of *Micrococcus leteus* cell suspensions at 450 nm was very similar for lysozyme from the two sources, while buffer alone did not have any effect on the reduction of turbidity.

Three selected suspension cell culture lines were induced to express lysozyme and the yield estimated in parallel by ELISA and the enzymatic activity assay described above (Table 2). T-test analysis showed that there was no significant difference between the lysozyme concentration measured by ELISA and enzymatic activity assay ($p < 0.05$). These results demonstrate that active recombinant human milk lysozyme is synthesized and maintained in rice callus cells and can be isolated without losing its activity.

Table 2. Comparison of Lysozyme Yields Estimates by Enzymatic Activity Assay and ELISA

Cell line	Lysozyme yield by enzymatic activity assay (lysozyme/total protein µg/mg)	Lysozyme yield by ELISA (lysozyme/total protein µg/mg)
156-5	25.8 +/- 6.3	30.3 +/- 3.9
156-16	32.1 +/- 5.7	32.9 +/- 3.2
156-31	47.0 +/- 6.2	42.3 +/- 7.0

F. Recombinant Human Lysozyme Has Bactericidal Function

The sensitive lysis of *Micrococcus luteus* cells in a turbidimetric assay (Fig. 4) indicates that recombinant human lysozyme possesses enzymatic activity and functions as a bactericide. To confirm this with a gram-negative bacterium, a bactericidal assay was carried out using an *E.coli* strain (JM109) as a test organism (Fig. 3).

In carrying out the assay, an aliquot of overnight JM109 culture was grown in LB medium until mid log phase. A standard innoculum of mid-log phase JM109 at 2 x 10^5 CFU (colony forming units)/ml was used in the bactericidal assay. Buffer (20 mM Sodium phosphate, pH 7.0, 0.5 mM EDTA) alone, buffer containing human milk lysozyme or rice seed derived lysozyme at about 30 µg/ml were sterilized by filtration. The mixture of cells and lysozyme solution was then incubated at 37° C for the specified length of time. One-fifth of the mixture volume was plated onto the LB agar plates and incubated overnight at 37 C in order to determine the number of colony forming units. At the concentration of 30 µg/ml, recombinant human lysozyme exhibited a similar bactericidal effect as lysozyme from human milk. There was no reduction of colony forming units using an extract from the non-transgenic control.

G. Purification of Lysozyme From Rice Calli, Suspension Cultures And Transgenic Rice Grains

Five rice calli lines expressing high levels of lysozyme were propagated and induced by sucrose starvation. The calli or cells were ground by a Tissuemizer in extraction buffer (PBS, 0.35 M NaCl) at 2 ml buffer/g of wet calli. The resulting tissue homogenate was centrifuged at 25,000 x g for 30 minutes at 4 C. The supernatant was removed and subjected to filtration through a pre-filter and then through a 0.45 µm nitrocellulose filter.

Approximately 1 liter of filtered supernatant from 500 grams of induced wet calli were then dialyzed against 50 mM sodium phosphate, pH 8.5 at 4 C overnight. The supernatant was loaded onto a 200 ml SP Sepharose fast flow column (XK26/40, Pharmacia) equilibrated with the loading buffer (50 mM sodium phosphate, pH 8.5) at a flow rate of four ml/min. The column was then washed with the same buffer until a baseline of A280 was achieved. Lysozyme was eluted by 0.2 M NaCl in the loading buffer and fractions containing lysozyme activity were pooled, concentrated and reapplied to a Sephacryl-100 column equilibrated and run with PBS at a flow rate of one ml/min. Proteins were eluted and separated by using PBS at a flow rate of one ml/min. Pure lysozyme fractions were identified by activity assay and total protein assay (Bradford) and the purity of lysozyme was confirmed by SDS-PAGE.

The five lines with the highest lysozyme expression level were selected and propagated continuously in petri dishes or shake flasks for lysozyme isolation and purification. A crude extract from rice callus contains both recombinant human lysozyme and large amounts of native rice proteins. Since the calculated pI of lysozyme is

approximately 11, a strong cation exchange column, SP-Sepharose fast flow (Pharmacia), was chosen as the first column to separate the rice proteins from recombinant human lysozyme. Most of the rice proteins did not bind to the column when equilibrated with 50 mM sodium phosphate, pH 8.5. The recombinant human lysozyme, on the other hand, bound to the column and was eluted by 0.2 M NaCl. Rice proteins that co-eluted with recombinant human lysozyme, were separated from lysozyme by gel filtration through a Sephacryl S-100 column and highly purified recombinant human lysozyme was obtained.

To purify human lysozyme from rice grains, R_2 rice seeds from transgenic plants were dehusked and milled to flour using conventional methods. Lysozyme was extracted by mixing the rice flour with 0.35 N NaCl in PBS at 100 grams/liter at room temperature for one hour. The resulting mixture was subjected to filtration through 3 µm of a pleated capsule, then through 1.2 µm of a serum capsule and finally through a Suporcap 50 capsule with a 0.8 µm glass filter on top of 0.45 µm filter (Pall, MI).

The clear rice extract (1 liter) was then dialyzed against 50 mm sodium phosphate, pH 8.5 at 4°C overnight and the dialyzed sample was loaded onto a cation exchange resin SP-Sepharose (Pharmacia Amersham), which was pre-conditioned with 50 mm sodium phosphate, pH 8.5 before loading. After loading, the column was washed with the same buffer until a base line A280 reading was achieved, then lysozyme was eluted with 0.2 N NaCl in 50 mm sodium phosphate, pH 8.5. Fractions containing lysozyme were pooled and reapplied to a Sephacryl S-100 column (Bio-Rad; equilibrated and run with PBS). Pure lysozyme was fractions were identified by enzymatic assay and total protein assay (Bradford). Finally the purity of lysozyme was confirmed by SDS-PAGE.

H. Attributes Of Recombinant Human Lysozyme Produced In Rice

(i). N-Terminal Amino Acid Sequencing

Recombinant human lysozyme (rLys) isolated from rice cells as described above, was separated by 18% SDS-PAGE followed by electroblotting to a PVDF membrane (Bio-Rad, CA). The lysozyme band was identified by staining the membrane with 0.1% Coomassie Brilliant Blue R-250 in 40% methanol and 1% glacial acetic acid for 1 minute. The stained PVDF membrane was immediately destained in 50% methanol until the band was clearly visible. After the blot was thoroughly washed with H_2O and air-dried, it was sequenced with a sequencer ABI 477 by Edman degradation chemistry at the Protein Structure Laboratory of the University of California at Davis. The results showed

that the rLys produced in transgenic rice seed had an identical N-terminal sequences to the human lysozyme, as follows:

Recombinant Lys----------LysVaLPheGluArg( )GluLeuAlaArgThr

Human Lys----------LysValPheGluArgCysGluLeuAlaArgThr

The blank parenthesis in recombinant lysozyme represents residue Cys which cannot be detected by the machine. This cycle was not defined, and could be due to the un-modified cysteine residue which cannot form a stable derivative in Edman degradation analysis.

Additionally, a number of structural and functional attributes of human lysozyme and recombinant lysozyme produced in rice were found to be the same, including molecular weight, pI, bactericidal effect with E. coli, thermal and pH stability and specific activity.

(ii). Thermal and pH stability of lysozyme

For biotechnological applications of the recombinant human lysozyme, its thermal and pH stability as well as its resistance to proteases is of decisive importance. A human lysozyme standard and lysozyme from rice were diluted to a final concentration of 50 µg/ml in PBS and subjected to the following thermal treatment in a sequential mode: (1): 62 °C for 15 minutes; (2): 72 C for 20 seconds; (3): 85 C for 3 minutes and finally; (4): 100 C for about 8 to about 20 seconds. Studies were conducted with 100 µl per tube and repeated three times. Aliquots were saved at the end of each treatment and the remaining lysozyme activity was measured by activity assay. The result showed that recombinant lysozyme exhibited the same degree of thermal stability in the temperature range from 62 °C to 100 °C as human lysozyme.

In another embodiment, approximately 50 µl of Hlys or rHlys was dissolved in PBS at 100 µg/ml and subjected to heat treatment. Four different temperatures of 65°C, 72°C, 85°C and 100°C were tested. With each temperature, 0 min, 0.33 min, 1.5 min, 3 min, 5 min and 15 min were selected to analyze the impact of incubation time on the stability of lysozyme (Fig 5A).

For studies on pH stability, lysozyme was dissolved in 0.9% NaCl at 100 µg/ml at pH 10, 9, 7.4, 5, 4, and 2. The solutions were incubated at 24 C for one hour. Experiments were conducted with 200 µl per tube and repeated three times. Remaining lysozyme was detected by lysozyme activity assay.

For pH treatments at pH 2, 4 and 5, Hlys and rHlys was dissolved in PBS adjusted to the corresponding pHs with HCl at 100 µg/ml. For pH 9 and 10, lysozyme was dissolved in TBS and 150 mM sodium carbonate/bicarbonate at 100 µg/ml,

respectively. Approximately 100 µl of lysozyme solution was incubated at 37°C for 30 min. The lysozyme activity was assessed by activity assay (Fig 5B).

Both Hlys and rHlys displayed similar thermal and pH stability.

(iii). Determination of *in vitro* protease resistance of lysozyme

Lysozyme was dissolved in 0.9% NaCl at 100 µg/ml. The pH of the solution was reduced to 3, 4 and 5 with HCl. Pepsin (Sigma, MO) (pepsin: lysozyme = 1:22 (w/w)) was added and the solutions were incubated at 37°C for one hour. Then the pH of all treatments was raised to pH 7 with bicarbonate. Pancreatin (Sigma, MO) (pancreatin : lysozyme = 1:110 (w/w)) was added to the neutral solution and incubated at 37 C for two hours. The remaining lysozyme activity was measured by activity assay.

In *in vitro* digestion experiments with pepsin and pancreatin, the native and recombinant human lysozyme displayed very similar resistance to pepsin and pancraetin digestion. Under these conditions, human albumin was degraded as demonstrated by SDS-PAGE (data not shown).

(iv). Biochemical Characterization of Lysozyme

After recombinant human lysozyme was purified to near homogeneity, several biochemical characterizations were carried out to compare human milk lysozyme with recombinant human milk lysozyme derived from rice cells. The results summarized in Table 3 show that by SDS-PAGE, native human milk lysozyme and recombinant lysozyme migrated to the same position.

Nucleotides encoding the rice *Ramy*3D signal peptide were attached to the human lysozyme gene in the expression vector pAPI156. Determination of the N-terminal amino acid sequence of the purified recombinant human lysozyme revealed an N-terminal sequence identical with that of native human lysozyme, as detailed above. Rice cells thus cleave the correct peptide bond to remove the *RAmy*3D signal peptide, when it is attached in the human lysozyme precursor.

The overall charge of recombinant and native human lysozyme were compared by isoelectric-focusing (IEF) gel electrophoresis and pI values determined. Since lysozyme is a basic protein with a calculated pI of 10.20, the pI comparison studies were carried out by reverse IEF gel electrophoresis. Recombinant and native human lysozyme displayed identical pI, indicating the same overall charge (data not shown).

Recombinant human lysozyme derived from transgenic rice had a specific activity similar to the native lysozyme (200,000 units/mg (Sigma, MO), whereas, lysozyme from chicken egg whites had the expected 3-4 fold lower specific activity (Sigma, MO) (Fig. 4).

Table 3. Comparison of Biochemical Characteristics of Human Milk Lysozyme and Recombinant Lysozyme

Lysozyme source	N-terminal sequence	Size (kDa)	Glycosylation	Specific activity (units/mg)	pI
Human milk	KVFER C ELART	14	No	201,526	10.2
rice	KVFER(-)*ELART	14	No	198,000	10.2

*This cycle was not defined, and could be due to the un-modified cysteine residue which cannot form a stable derivative in Edman degradation analysis.

The results described above demonstrate the ability to use rice cells as a production system to express human lysozyme from milk. Over 160 individual transformants were screened by immunoblot, enzymatic activity assay and ELISA. Yields of recombinant human milk lysozyme reached 4% of soluble cell proteins in culture cells and over 40% of soluble proteins in rice grains. Although the mechanism is not part of the invention, the high expression level may be explained by the utilization of the strong *RAmy*3D promoter (Huang *et al.*, 1993) in culture cell system and Gt1 promoter in grain expression system and the codon-optimized gene.

The plant derived human milk lysozyme obtained by the methods of the present invention was identical to endogenous human lysozyme in electrophoretic mobility, molecular weight, overall surface charges and specific bactericidal activity.

EXAMPLE 4

Characterization of Recombinant Human Lactoferrin (rLF) Produced by Transgenic Rice Plants

A. Southern blot analysis

About three grams of young leaf were collected and ground with liquid nitrogen into a very fine powder. The DNA was isolated according to the procedure as described in Dellaporta *et al.*, 1983, and purified by phenol-chloroform extraction. Approximately 5 µg of *ECo*RI and *Hin*dIII digested DNA from each line was used to make blot for Southern analysis. The ECLTM direct nucleic acid labeling and detection system (Amersham, USA) was used for analysis.

The lactoferrin gene copy number was estimated to be from about 1 to about 10 as determined by Southern blot hybridization using *Eco*RI and *Hin*dIII digested genomic DNA. The API164-12-1 (R0) transgenic plant line was subjected to Southern analysis

together with ten Western blot positive, field grown R1 lines. A typical Southern blot shows that there are at least three fragments above the original plasmid derived plant transformation unit (3156 bp). All the LF inserts appear to be inherited from the original R0 transgenic plant event to R5 generation.

B. Protein Isolation and Western blot

Rice seeds were ground with 1 ml of 0.35 N NaCl in phosphate buffer saline (PBS), pH 7.4 using an ice-cold mortar and pestle and the resulting homogenate was centrifuged at 15000 rpm for 15 min at 4°C. The supernatant was used as a protein extract and about 1/25 or 1/50 of the salt soluble content was loaded onto a 10% pre cast gel (Novex, USA) and electrophoresis was carried according to the manufacturer's instructions. For total protein detection, the polyacrylamide gel was stained with 0.1% Coomassie brilliant blue R-250 (dissolved in 45% methanol and 10% glacial acetic acid) for at least three hours and destained with 45% methanol and 10% glacial acetic acid until the desired background was achieved.

For Western blot analysis, SDS-PAGE gels were electroblotted onto a 0.45 µm nitrocellulose membrane with a Mini-Trans-Blot Electrophoretic Transfer Cell System (Bio-Rad, USA) and subsequently subjected to immuno-blotting analysis. The blot was blocked with 5% non-fat dry milk in PBS for at least two hours followed by three washes with PBS for 10 minutes each. The primary rabbit polyclonal antibody against hLF (Daka A/S, Denmark) was diluted at 1:2500 in the blocking buffer and the blot was incubated in the solution for one hour. The blot was washed with PBS for three times with 10 minutes each. The secondary goat anti-rabbit IgG (H+L)-alkaline phosphatase conjugated (Bio-Rad, USA) was diluted in the blocking buffer at 1:5000 ratio. The membrane was incubated in the secondary antibody solution for one hour and followed by three washes with PBS. Color development was initiated by adding the substrate system BCIP-NBT (Sigma, USA) and the process was stopped by rinsing the blot with H2O once the desirable intensity of the bands was achieved.

One hundred eight (108) R0 plants were grown to maturity, seeds were harvested from 56 fertile plants and individual seeds analyzed by Western blot to detect the expression of rLF. Coomassie blue staining was carried out to compare the mobility of rLF with native human lactoferrin (hLF) (Fig. 8), with 40 µg of total protein loaded onto each lane, along with 40 ng of native purified hLF per lane as the positive control.

Estimation of total rLF by ELISA indicated that from 93 µg to 130 µg rLF was expressed in transformed rice seeds. A typical Western blot analysis (Fig. 9) illustrates

that both rLF and native hLF migrate at approximately the same rate with the molecular weight about 80 kDa, consistent with that determined by other researchers (Wang *et al.*, 1984).

C. Protein purification

Rice seeds from R2 homozygous generation were dehusked and milled to flour conventionally. Recombinant lactoferrin was extracted by mixing the rice flour with 0.35 N NaCl in PBS at 100 g/l at room temperature for two hours. The resulting mixture was centrifuged at 15,000 rpm for one hour at 4°C. The collected supernatant was subjected to the following steps of filtration before loading onto a Sepharose column. First, the supernatant was run through a few layers of cheesecloth. Then the filtrate was passed sequentially through an 8µm paper, 1 µm paper and a 0.25µm nitrocellulose membrane. The clear protein solution was loaded onto a ConA Sepharose column (Pharmacia, XK 26) which had been equilibrated with 0.5 N NaCl in 20 mM Tris, pH 7.4 (binding buffer) at a flow rate at 4 ml/min. After the loading was complete, the column was washed with binding buffer until the baseline at A280nm was achieved. Lactoferrin was eluted with 0.1N mannoside in the binding buffer. Fractions containing lactoferrin were pooled and loaded onto a second column SP-Sepharose (Bio-Rad, USA) which has been equilibrated with 0.4 N NaCl in 50 mM sodium phosphate, pH 8.0 (binding buffer B) at the flow rate 4 ml/min. Then the column was washed with the binding buffer B until the baseline at A280 nm was obtained. Lactoferrin was eluted by 1 N NaCl in 50 mM sodium phosphate, pH 8.0 and the fractions containing LF were pooled and dialyzed against PBS. Finally the purity of LF was assessed by SDS-PAGE and stored at - 80°C.

In another embodiment, recombinant human lactoferrin (rHLF) was extracted by mixing rice flour with 0.35 M NaCl in PBS at 75 g/L at room temperature for 2.5 hours. The extract was passed through six layers of cheesecloth before centrifugation (10,000 g for 1 hour at 4°C). The supernatant was recovered and the NaCl concentration was adjusted to 0.4 M (pH 8.0). After a second centrifugation at 10,000 g for 10 minutes at 4°C, the supernatant was collected and filtered through 0.45 µm nitrocellulose membrane. The filtrate was loaded onto a SP-Sepharose column (Bio-Rad, Hercules, CA) which had been equilibrated with 0.4 M NaCl in 50 mM sodium phosphate, pH 8.0 (binding buffer) at a flow rate of 4 ml/min. The column was washed with the binding buffer until baseline A280 was obtained. Lactoferrin was eluted by a linear gradient and dialyzed against PBS. The purified rHLF was analyzed by SDS-PAGE and stored at – 80°C.

D. Enzyme Linked Immunosorbant Assay (ELISA)

ELISA was conducted using seed extracts, isolated as described above, with total protein assayed using the Bradford method (Bradford, M., 1976). The ELISA was based on a typical sandwich format generally known in the art. Briefly, 96 well plates were coated with rabbit anti-human lactoferrin antibody (Daka A/S, Denmark), then rLF and control samples were added to individual wells of the plate and incubated for 1 hour at 35°C. Rabbit anti-human lactoferrin horseradish peroxidase conjugate (Biodesign, USA) was then added to each well and incubated for 1 hour at 35°C, followed by addition of the tetramethylbenzidine substrate (Sigma, USA) and incubation for 3 minutes at room temperature. The reaction was stopped by adding 1N H2SO4 to each well. The plates were read at dual wavelengths of 450 and 650 nm in a Microplate Reader (Bio-Rad, model 3550) and the data was processed by using Microplate Manager III (Bio-Rad). The results of an analysis of 10 homozygous selected lines showed that from 93 µg to 130 µg rLF was expressed per seed.

E. Selection of plants for advance generations

At least 20 - 40 seeds from 11 independent lines were analyzed. Individual R1 seeds were cut into half and endospermic halves were subjected to analysis by Western blot with the positive corresponding embryonic halves germinated on 3% sucrose medium with 0.7% agar. The seedlings were transplanted to the field for R1 generation. Out of 11 individual lines, 3 lines were expressed. A total of 38 plants were grown in the field derived from the 3 expressed mother lines. Based on the agronomic character (Table 4) of those 38 plants, 28 plants were selected.

It was observed that all the Western positive R1 seeds were opaque to pinkish in color in comparison to control seeds, so this criterion was applied in screening the R2 seeds. Mature R2 seeds were harvested at maturity and dehusked. The pinkish R2 seeds were confirmed by Western dot blot and ELISA as expressing rLF (data not presented). Finally 10 homozygous R2 lines were selected and grown in the field in order to advance the generation.

Table 4. Comparison Of Phenotypic Characteristics Of Native TP-309 And Transformed TP-309 Rice Seeds

Source	Effective tiller	Blank grain (%)	1000 seed weight (g)	µg of rLF/seed
TP-309	43	5.0	25	
Homozygous transgenic lines	42	19.7	20.2	125

During R2 and R3 generation the percentage of blank seeds was higher in homozygous transgenic lines than in the non-transgenic control. This affected the 1000 seed weight. However, in the R4 generation no significant differences in phenotypic character were observed in homozygous transgenic lines when compared to non-transformed TP309 (Table 4).

F. Attributes Of Recombinant Human Lactoferrin Produced In Rice

Physical characterization of the rLF showed there was no significant difference between the rLF and a commercially available purified form of hLF based on N-terminal amino acid sequencing, and physical characteristics of rLF such as molecular weight as determined by MALDI-MS, HPLC profile of which showed a comparable peptide map, pH dependent iron release and bacteriostatic activity, using the analyses described below.

(i). N-Terminal Amino Acid Sequencing

Purified rLF from rice seeds was resolved by 10% SDS-PAGE, followed by electroblotting to PVDF membrane (Bio-Rad, USA). The target band was identified by staining the membrane with 0.1% Coomassie brilliant blue R-250 in 40% methanol and 1% glacial acetic acid for 1 minute. The stained PVDF membrane was immediately destained in 50% methanol until the band is clearly visible. The blot was thoroughly washed with ddH2O and air dried. Finally this sample was sent to the Protein Structure Laboratory in University of California at Davis (CA, USA) for sequencing analysis.

(ii). Detection of glycosylation and determination of sugar content

Glycosylation of the recombinant human lactoferrin produced in rice was analyzed by an immunoblot kit for glycoprotein detection (Bio-Rad, USA) per instructions from the manufacturer. An increase of molecular weight of lactoferrin due to carbohydrate content was determined by Matrix Assisted Laser Desorption Ionization-Mass spectrometry (MALDI-MS) (PE Applied Biosystems, Voyager System).

Recombinant lactoferrin produced in rice is glycosylated as evident from the binding to Con A resin, the positive staining by glycoprotein detection kit as well as the larger detected mass as compared to the calculated mass (76.2 kDa) based on the peptide backbone. MALDI-MS showed that seed derived recombinant lactoferrin has molecular weight of 78.5 kD while human milk lactoferrin is 80.6 kDa (Table 5). The difference could be due to the lesser degree of glycosylation in the rice seed-derived lactoferrin. Analysis shows that the purified rHLF contains xylose but lacks sialic acid, which is consistent with plant post-translational modification patterns (Matsumoto *et al.*, 1995).

(iii). Determination of isoelectric point of Lactoferrin

Reverse isoelectric focusing (IEF) gel electrophoresis was carried out with a precast Novex IEF gel, pH 3-10 according to the manufacturer' instruction. About 30 µg of purified rLF was loaded and the running condition was 100 V for 50 minutes and 200 V for 20 minutes. The gel was then fixed in 136 mM sulphosalicylic acid and 11.5% TCA for 30 minutes, stained in 0.1% Coomassie brilliant blue R-250, 40% ethanol, 10% glacial acetic acid for 30 minutes and destained in a solution containing 25% ethanol and 8% acetic acid.

(iv). Comparison of physical characteristics of rLF with native hLF

The HPLC profile of native and rLF showed a comparable peptide map. This confirmed that LF from the two sources have an identical amino acid sequence (data not presented). Additional comparisons confirm that human lactoferrrin produced in transgenic rice closely resembles native human lactoferrin, as evidenced by (1) the N-terminal sequence of purified rLF from homozygous R2 seeds and hLF (Dakao A/S, Denmark), which were shown to be identical (Table 5); (2) the isoelectric point (pI) of native and rice seed derived LF which is the same, indicating that they have similar surface charges (Table 5); (3) the pH dependent iron release of rLF which was shown to be closely related to that of native hLF (Fig. 11 and see section vii of example 4); and (4) the bacteriostatic activity of rHLf which was shown to be similar to that of native human lactoferrin (nHLf) on enteropathogenic *E. coli* (EPEC; Fig. 10) and confirmed the presence of active recombinant LF in extracts derived from transformed rice seeds (see section ix of Example 4).

Table 5. Physical characterization data for human (hLF) and rice seed derived recombinant lactoferrin (rLF)

LF source	Size (kDa)	N-terminal sequence	pI	Glycosy-lated	Sugar content (%)
hLF	80.6	GlyArgArgArgArgSerValGlnTrpCysAla	8.2	YES	5.5
rLF	78.5	GlyArgArgArgArgSerValGlnTrp( )Ala	8.2	YES	2.9

(v). Iron content and nutrient value determination of rice seeds

The iron content of R2 homozygous seeds was determined. Two grams of dry mature seeds from each transformed and non-transformed line were weighed and wet-ashed with HNO3 and H2O2 solution at 110°C (Goto *et al.*, 1999). The ash was dissolved in 1N HCl solution. The iron content was then measured by absorbance of Fe-O-phenanthrolin at 510 nm, using a Sigma kit (Sigma, USA) per instructions of manufacturer.

The different values of nutrient facts of homozygous transgenic seeds and non transgenic seeds were measured by standard procedure at A & L Western Agricultural Laboratories (Modesto, CA, USA).

A comparative analysis of transgenic lactoferrrin-expressing rice seeds with non transformed native Teipei-309 showed that there is no significant difference between transformed and non transformed seeds in nutrient value with the exception that the concentration of iron is 50% greater (Table 6). The increased level of iron may be the reason for the opaqueness and pink coloration of the rLF expressing transgenic rice seeds.

In another embodiment, 0.2 grams of dried, dehusked grains expressing rHLF were wet-ashed with concentrated HNO3 for two days and dissolved in 5 ml of DDI H2O. The iron contents of the samples were measured by flame atomic absorption spectrophotometry (Thermo Jarrel Ash SH4000, Franklin, MA). NIST liver was analyzed concurrently to verify the accuracy of the standard curve.

The iron content of transgenic rice grains was more than twice that of non-transformed TP309 grains, while there were no significant differences in other tested nutrition factors between transformed and non-transformed grains (Table 7). This suggests that groups ingesting transgenic rice with rHLF will increase the iron intake.

The transgenic grains with increased iron content were opaque-pinkish in color. The opaque-pinkish color was observed inside as well as outside the rice endosperm. This opaque-pinkish color, segregated in Mendelian fashion, was linked with expression of rHLF and was inherited through the R4 generation.

There was no difference noticed during the seed germination of transgenic seeds, the phenotype of R2 R3 and R4 plants was vigorous and the seed yield was similar to that of non-transgenic Teipei-309 plants (data not shown).

Table 6. Comparison of Nutrition Value (in mg) Per 100 Gram of Non Transformed and Transformed Rice Seeds

Source	Carbohy-drate	Protein	Fat	Ca	K	Na	Fe	Water	Calories
TP-309	76.0	8.7	2.4	9	370	<10	0.8	11.3	369
Homozygous transformed lines	75.7	8.7	2.2	8	330	<10	1.2	11.8	367

Table 7. Comparison of Mineral Contents (in µg) Per Gram of rHLF-Transformed and Non-Transformed Rice Grains

Source	Cu	Fe	Mn	Zn
Non-transformed	2.9	8.7	33.1	20.8
Transformed	4.7	19.2	17.7	28.7

(vi). Tissue specificity and stability of rLF

An endosperm specific rice glutelin promoter was used to express recombinant lactoferrin in maturing or matured seeds. To confirm the tissue specificity of the expressed lactoferrin, protein was extracted from root, shoot, leaf beside mature seed and subjected to Western blot and the results indicated that there was no detectable expression of rLF except in the seed/endosperm (Fig. 9). Furthermore, the presence of rLF in 5 day old germinated seeds showed the stability of stored rLF within the plant cell during germination.

(vii). Iron Saturation And pH Dependent Iron Release

Lactoferrin was incubated with 2M excess ferric iron ($FeCl_3$:NTA = 1:4) and sodium bicarbonate (Fe:HCO_3^- = 1:1) for 2h at room temperature. Excess free iron was removed by using a PD-10 desalting column (Pharmacia, USA) and the iron saturation level was determined by the A280/A456 ratio. Both native hLF and rLF were completely saturated by iron. Holo hLF was incubated in buffers with a pH between 2 and 7.4, at room temperature for 24 h. Free iron released from hLF was removed and the iron saturation level was determined by A280/A456 ratio.

The results showed that iron release was similar for both hLF and rLF. Iron release began around pH 4 and was completed around pH 2 (Fig. 11). The iron binding was reversible since iron-desaturated rLF was re-saturated by raising the pH to 7 (data not shown). The similarity in pH dependent iron release of rLF to that of the hLF standard demonstrated that rLF is able to adapt the appropriate tertiary structure for proper iron binding and release (Salmon, Legrand *et al.* 1997).

(viii). Binding and uptake by Caco-2 cells

50,000 Caco-2 cells/well were seeded and grown in Minimum Essential Medium (GIBCO, Rockville, MD) containing 10% fetal bovine serum in 24 or 48 well tissue culture plates for 3 weeks. For binding studies, Caco-2 cells were incubated with varying concentrations (0-2 μM) of 125I-HLf in the presence or absence of 100-fold excess of unlabeled nHLf for 2 hours at 4 oC and cells were washed 5 times with ice-cold PBS. Cells were solubilized with 0.5 ml of 0.1% SDS and radioactivity was quantified in a gamma counter. For uptake studies, 0.4 μM of 125I-HLf was incubated with Caco-2 cells for 0 to 24 hours at 37 °C and cells were washed, dissociated by the same way as in the binding study. 0.5 ml of 24% TCA solution was added to the dissociated cells and free iodine was removed by the centrifugation. Free and protein-bound ^{125}I were quantified separately to evaluate how much of HLf was degraded in the cells. Receptor-binding of rHLf to the human intestinal Caco-2 cell line was saturable and specific, indicating that rHLf bound to the Lf receptor. The binding constant was similar for rHLf and nHLf, but the number of binding sites was slightly higher for rHLf, which may be due to the difference in glycosylation. Uptake of HLf by Caco-2 cells was identical for rHLf and nHLf.

(ix). *In vitro* Digestion: effect on antimicrobial activity and binding/uptake to Caco-2 cells

Lactoferrin is known to inhibit the growth of a variety of bacterial species based on its iron chelation and direct bactericidal properties. The anti-microbial effect of rLF extracted from rice seeds was tested following treatment using an *in vitro* digestion model with an enzymatic system containing pepsin (an enzyme active in stomach) and pancreatin (an enzyme active in deodenum).

LF proteins were dissolved in PBS at 1mg/ml, and either left untreated, pepsin treated (0.08mg/ml at 37°C for 30 min), or pepsin/pancreatin treated (0.016 mg/ml at 37°C for 30 min). LF proteins were sterilized by passing through a membrane filter with a pore size of 0.2 μm [Rudloff, 1992]. The filter sterilized LF (0.5μg/ml) was incubated with 104 colony forming unit (CFU) enteropathogenic *E. coli* (EPEC)/μl in 100 μl sterile

synthetic broth (1.7% : AOAC) containing 0.1% dextrose and 0.4 ppm ferrous sulfate at 37°C for 12h and colony forming units (CFU) were determined.

Starting with an enteropathogenic *E. coli* (EPEC) concentration of 10^4 CFU (colony forming units), the untreated samples of rLF reached up to $10^{6.5}$ CFU after 12 h of incubation at 37°C in comparison to hLF, which produced up to 10^6 CFU. An *in vitro* digestion model using an enzymatic system containing pepsin (enzyme active in stomach) and pancreatin (enzyme active in deodenum) with moderate shaking to imitate the transit of protein through infant gut [Rudloff, 1992] was used. rLf and nHLf were treated with active pepsin and pancreatic enzymes and exposed to 10^4 CFU EPEC cells for 12 h at 37°C (Fig. 10). Both the native human lactoferrin standard (nHLf) and the recombinant rice-derived lactoferrin (rLf) remained active in inhibiting growth of enteropathogenic *E. coli,* indicating that both nHLf and rHLf are resistant to protease digestion.

SDS-PAGE and ELISA revealed that nHLf and rHLf resist digestion by pepsin (at pH 3.8) and pancreatin, whereas human serum albumin is completely digested after *in vitro* digestion. Western blots revealed that immunoreactivity was also maintained after digestion. Although some smaller molecules were generated during digestion of HLf, most of the immunologically detectable HLf retained its intact size. More than 50 % of rHLf and nHLf was immunologically detectable by ELISA, but ^{125}I-HLf was around 40 % and ^{59}Fe-HLf was only 20 % detectable, indicating that ELISA detects small peptide fragments of HLf, which are removed by the PD-10 column and that about 50-60 % of Fe was released from detectable HLf after *in vitro* digestion. The iron-holding capacity was not significantly different.

The dissociation constant (Kd) and the number of binding sites for HLf to its receptor were determined from the binding study. Both Kd and the number of bindings sites were not significantly different between nHLf and rHLf after *in vitro* digestion (Figure 12A, 12B). Digestion did not appear to affect on the Kd but made the number of binding sites much lower. Total Lf uptake was not significantly different between nHLf and rHLf after *in vitro* digestion (Figure 12C), though uptake was about one third when compared with undigested nHLf. Total iron uptake from nHLf was twice as high as that from rHLf. Percent degradation of HLf was similar regardless of digestion or not, and the native or recombinant form (Figure 12D).

(x). Thermal Stability: effect on antimicrobial activity and binding/uptake to Caco-2 cells

1.0 mg/ml of holo-HLf in PBS was treated by the following conditions: (a) 62 °C for 15 minutes, (b) 72 °C for 20 seconds, (c) 85 °C for 3 minutes, or (d) 100 oC for 8 seconds. Survival ratio of HLf determined by ELISA were more than 90% following treatment at 62 °C for 15 minutes, at 72 °C for 20 seconds, or at 85 °C for 3 minutes, but it was considerably lower after 100 °C for 8 seconds. This high temperature precipitated both types of HLf and only 10% of HLf was detectable by ELISA. More than 80% of iron was still bound to both rHLf and nHLf after all thermal treatments with the exception of 100 °C for 8 sec. In 10% of survived HLf after 100 °C for 8 sec, the iron saturation level of nHLf was above 80% whereas that of rHLf was only about 40%.

SDS-PAGE and Western blots revealed no difference in immunoreactivity between nHLf and rHLf at 62 °C for 15 minutes, at 72 °C for 20 seconds, and at 85 °C for 3 minutes, but at 100 °C for 8 seconds, rHLf almost completely lost its immunological activity, whereas nHLf still maintained detectable immunoreactivity.

There was no significant difference in anti-microbial activity between nHLf and rHLf after heat-treatment. Anti-microbial activity of HLf was not affected by treatment at either 62 °C for 15 min, 72 °C for 20 sec or 85 °C for 3 min.

The Kd and the number of binding site for nHLf and rHLf were not significantly different at 62 °C and 72 °C though there is a trend that nHLf is somewhat lower Kd and binding sites than rHLf. As the temperature was increased (such as 85 °C and 100 °C), more rHLf bound to Caco-2 cells, most likely by non-specific binding due to more rHLf being denatured than nHLf. Uptake properties were similar for nHLf and rHLf even in the group treated at 100 °C where uptake of both types of HLf was highest among all the thermal treatments. Free iodine levels in the cells were also evaluated since it reflects degradation of HLf. About 20% of HLf was degraded in the untreated sample. There was no significant difference between nHLf and rHLf. Interestingly, samples treated at 100 °C were degraded twice as much as untreated samples of nHLf and rHLf, which may indicate that denaturation of HLf caused by heat treatment will make the protein more susceptible to proteases in the cells.

(xi). <u>pH Stability: effect on antimicrobial activity and binding/uptake to Caco-2 cells</u>

1.0 mg/ml of holo-HLf in PBS was adjusted to pH 2, 4, 6, or 7.4 by the addition of 1 M HCl and incubated for 1 h at room temperature. The pH was then adjusted to 7.0 with 1 M NaHCO3. Free iron released from HLf, was removed by a desalting column.

After low pH treatment, 100% of both nHLf and rHLf survived. The iron-holding capacity was maintained in all samples and the iron saturation level was above 95%. SDS-PAGE and Western blots revealed that there was no difference between nHLf and rHLf for any of the treatments. A slightly smaller immunoreactive molecule (~70 kD) was detected after exposure of nHLf to pH 2 and 4 and of rHLf to pH 2.

Antimicrobial activities of nHLf and rHLf were stable after exposure to low pH in the range of pH 2.0 to 7.4. As the pH was lowered, the activity of rHLf appeared to be higher and constant, whereas nHLf did not show any pH dependency.

Kd and the number of binding sites for nHLf were not significantly different from those for rHLf but a trend was always lower for nHLf within the range of pH 2.0 to 7.4, which is similar to control and thermal treatment samples. The Kd and the number of binding sites for nHLf and rHLf were unaffected by pH treatment down to 2.0 for 1 hour. Uptake properties were similar for nHLf and rHLf in the pH range of 2.0 to 7.4. Degradation of HLf in Caco-2 cells was also evaluated and there was no significant difference between nHLf and rHLf.

EXAMPLE 5

Generation and Characterization of Recombinant Human α-1-antitrypsin (AAT) Produced by Transgenic Rice Plants

A. Construction and expression of human AAT in rice cells

The construction and purification of functional recombinant human AAT were carried out as exemplified in previous examples. Briefly, codon-optimized AAT gene was cloned into an pAPI145 that contains the rice Gt1 promoter, Gt1 signal peptide, and Nos terminator, pAPI241 that contains Glb promoter, Glb signal peptide, and Nos terminator, and API280 that contains Bx7 promoter, Bx7 signal peptide, and Nos terminator, as exemplified in Example 1. The resulting plasmids were named pAPI250, API255 and pAPI282, respectively (Figure 13). Transgenic plants expressing AAT were generated as above, and plant-generated recombinant AAT was characterized. To express AAT in culture cells, codon-optimized AAT gene was cloned into an expression cassette that contains the rice RAmy3D promoter, signal peptide, and terminator. Recombinant AAT expression was induced and secreted to the culture medium under the sugar starvation condition. Purification of rAAT was achieved through a scheme that consisted of an affinity column (Con A), anion exchange column (DEAE), and a hydrophobic interaction column (Octyl).

B. SDS- Polyacrylamide Gel Electrophoresis (SDS-PAGE)

AAT samples were ground with PBS with mortar and pestle. The resulting extract was spun and 20 microliters of supernatant loaded into a precast SDS-PAGE gel. The AAT protein was clearly visualized with Coomassie brilliant blue staining (Figure 14).

C. Western Blot Analysis

For immunoblotting analysis, gels were electroblotted to a 0.45 µm nitrocellulose membrane with a Mini Trans-blot Electrophoretic Transfer cell (Bio-Rad, USA) and subsequently subjected to immunoblotting analysis. Blots were blocked with 5% non-fat dry milk in PBS, pH 7.4 for at least two hours followed by three washes with PBS, pH 7.4 for 10 minutes each. The primary rabbit polyclonal antibody against human alpha-1-antitrypsin (Dako A/S, Denmark) was diluted to 1:2500 in the blocking buffer and the blot was incubated for at least one hour. The blot was then washed as described previously. The secondary antibody, goat anti-rabbit IgG (H+L)-alkaline phosphatase conjugated (Bio-Rad), was diluted in the blocking buffer at a dilution of 1:4000. The membrane was then incubated in the secondary antibody solution for one hour and followed by the same wash process. Color development was initiated by adding the substrate BCIP/NBT from Sigma.

The western result showed that AAT protein is clearly visualized and confirmed that AAT expressed and deposited in transgenic rice grain, has a molecular weight that is somewhat smaller than that of native AAT (Figure 15).

D. ELISA

Standards for this assay ranged from 1.25 - 20 ng/mL of AAT (Athena) diluted in PBST. Nunc Immuno-plate Maxisorp 96-well plates (Nunc, Denmark) were coated for 16 h at 4°C by a 1:10,000 dilution of rabbit anti-human AAT in 0.05 M sodium bicarbonate, pH 9.6. The plates were washed 3 times with PBST (PBS, pH=7.4, 0.05% Tween-20) and subsequently incubated with sample for 1 h at room temperature while rocking. The plates were washed again 3 times with PBST, followed by incubation with a 1:50,000 dilution of goat anti-human AAT conjugated to HRP for 1 h at room temperature. The plates were washed 3 times with PBST, and bound antibody was detected by the HRP/ hydrogen peroxide catalyzed reaction of TMB. The reaction was stopped with 2 M sulfuric acid, and the plates were read on a microtiter plate reader at 450 nm, using 620 nm as a reference filter.

Recombinant AAT is 2.1 times more immunoreactive, when comparing equal concentrations as determined by the Lowry assay.

E. AAT activity assay

AAT activity was analyzed using a modified method published by Travis and Johnson (1981). In 96-well microtiter plates, 60 µL samples diluted in Tris buffer, (0.2 M Tris, pH 8.0) were added. In each well, 60 µL of elastase [0.01 mg/mL porcine pancreatic elastase (PPE) in Tris buffer] was also added. The plate was rocked for 5 min at room temperature to allow any available AAT to bind to the elastase. Another 120 µL of substrate solution (10 M N-Succinyl-AAA-p-nitroanilide in DMSO diluted in Tris buffer to give 0.33 M N-Succinyl-AAA-p-nitroanilide) was added, and the plate was rocked for 1-2 min at room temperature. The plate was immediately read on a microtiter plate at 405 nm. The plate was read again after 5 min, and the change in absorbance was calculated. AAT activity was determined using linear regression from a standard curve. The results show that AAT protein produced in rice grain has similar bioactivity as that of native AAT.

F. Band shift assay

The unique property of the covalent-linked complex formed between AAT and PPE permits an analysis of the activity of AAT by SDS-PAGE. Briefly, 20 µl of tested samples containing AAT from the screening or purification processes was incubated with 100 ng PPE at 37°C for 15 minutes. Five µl of SDS-loading dye was added and the reaction mixture boiled for five minutes. The sample was then centrifuged and kept on ice until loaded onto a 10% precast SDS-PAGE gel. The resulting gel was stained with 0.1% Coomassie brilliant blue R-250 as described below. For immunodetection, a western blot analysis was carried as described above. Again the band shift assay indicated that AAT protein produced in rice grain has similar bioactivity as that of native AAT (Fig. 16A and 16B).

G. *In vitro* digestion.

The digestion was carried out using a modified method of Rudloff and Lonnerdal (1992) was used after some modifications. Native and recombinant AAT were diluted in PBS or formula to 0.5 mg/mL. Hydrochloric acid (1 M) was added to all samples to adjust the pH 3, 4, and 5, then 2.5 µL of 2% pepsin in 0.01 M HCl (3,100 U/mg solid) were added and all samples were placed in a shaking incubator for 30 or 60 min at

37°C. The pH was restored by drop-wise addition of 1 M NaHCO3, and 2.5 μL of 0.4% pancreatin in 0.1 M NaHCO3 were added. Samples were incubated for 1 or 2 hours at 37°C, and the reaction was halted by dilution 1:2 in sample buffer and boiling for 3 min. For samples subjected to pepsin digestion only, boiling was unnecessary since the pepsin was inactivated when the pH was raised above pH 6 with NaHCO3 (Piper and Fenton, 1965). The enzyme: substrate ratio was approximately 1:20 for samples in buffer only and about 1:600 for samples in formula.

A significant amount of recombinant and native AAT survived the *in vitro* digestion, and both forms were more resistant to degradation than human serum albumin. Digestion with pepsin at pH 4 shows that 65% of recombinant AAT is detectable by ELISA after digestion, which is similar to 67% of native AAT surviving. The trypsin assay shows that much of the inhibitory properties of both forms are still intact, and the activity assay reveals that 63% and 59% of the activity of native and recombinant AAT remains, respectively. When exposed to both pepsin and pancreatin in buffer, native AAT resisted degradation when the pH of the pepsin incubation was pH 4 or higher. Under this condition, the recombinant form was less resistant, although a large part remained after pepsin digestion at pH 5 and pancreatin digestion. At pH 4, more of the recombinant protein was degraded, either due to pepsin activity or pH instability. AAT activity could not be determined after digestion by pepsin and pancreatin because of the inactivation of pancreatin by boiling which also inactivates AAT activity. In formula, both forms appeared to be equally resistant to degradation. While both native and recombinant AAT were still present after pepsin digestion at pH 5 followed by pancreatin digestion, bands at about 33 kD (casein) are faint or missing. It is possible that other proteins in formula are preferentially cleaved, reducing the amount of AAT being digested.

H. Thermal Stability of recombinant human ATT

Both native and recombinant human AAT were diluted in phosphate buffered saline (PBS) or infant formula (Enfamil with Iron, Mead Johnson, Evansville, IL) to a concentration of 0.1 mg/mL. Samples, 100 μL in capped, 10 x 75 mm glass tubes, were treated as follows: 60°C for 15 min, 72°C for 20 sec, 85°C for 3 min, and 137°C (temperature of oil bath) for 20 sec. The samples were allowed to cool to room temperature after heat treatment. For formula samples with bile extract added, 2.5 μL of 12% porcine bile extract (Sigma) were added, then vortexed quickly, incubated at 37°C for 10 min, and vortexed again. All samples were diluted 1:10 in PBS and transferred to

1.5 mL tubes. Formula samples were centrifuged at 15,000g for 20 min to remove the insoluble fraction, and the supernatant was withdrawn after skimming off the fat. All samples were subsequently transferred to 1.5 mL tubes and analyzed.

The thermal stability of native AAT exceeded that of the recombinant form in buffer, but the recombinant AAT retained significant stability under most conditions. When heated in buffer only, SDS-PAGE and Western blots show that the two forms of AAT have similar structural stability. While the ELISA data show that the recombinant protein is less stable at the higher temperatures, the recombinant protein is similar to the native form under the other conditions. However, the functional stability of the recombinant protein may be affected. The thermal stability assay shows that the recombinant protein lost functional ability at several of the heat conditions, whereas the native protein was functional at all heat conditions except for at 62°C for 15 minutes. While the elastase-inhibiting properties of native AAT were about 90% after all heat treatments, 62 and 51% of the recombinant protein's activity remained after 85°C, 3 minutes, and 137°C, 20 seconds, respectively.

The heat treatments of native and recombinant AAT in formula affected the detection of the proteins, but the addition of bile extract following heat treatment restored antibody recognition of the recombinant form. While the Western blot data show less detectable protein only at 85°C, 3 min for the native AAT and at 72°C, 20 sec and 137°C, 20 sec for the recombinant AAT, the ELISA data shows less than 20% protein detected for both forms and for all heat conditions. When bile extracts were added to the heated formula samples, the ELISA data for the recombinant form showed that more than 50% was still detectable after heat treatment. The bile extract did affect detection of the native form by ELISA for most of the heat treatments. The Western blots corroborated the ELISA data and showed that the bile extract may dissociate the recombinant AAT from other formula proteins, but it is not effective for native AAT at the higher temperatures.

I. pH stability of human ATT

Native and recombinant AAT were diluted in PBS or formula to 0.1 mg/mL. The sample volume was 1 mL, and the pH of each sample was adjusted drop-wise with 1 M HCl. The range of pHs tested was from pH 2 to 8 for the samples in PBS and pH 2 to 7 for samples in formula. After a 1 hour incubation at room temperature, the pH was restored to pH 7 with 1M $NaHCO_3$. Formula samples were centrifuged as exemplified in above Thermal Stability section.

Both native and recombinant AAT appear resistant to low pH conditions in both PBS and formula. There were no differences between treatment groups and controls for pH 3 through 7, and controls or between the native and recombinant AAT according to SDS-PAGE, Western blots, and trypsin assay. However, the elastase assay and ELISA data show that recombinant AAT is more affected by acidic conditions than the native form. In PBS, native AAT was more than 95% intact, while about 60-80% of the recombinant AAT activity was intact. Infant formula may have a stabilizing effect on the recombinant protein, since it was found to be as stable as the native form according to ELISA and the Western blot.

Native and recombinant AAT can withstand acidic and digestive conditions as assessed by SDS-PAGE, Western blots, ELISA and activity assay. Native AAT regains much of its structural and functional stability after treatment at acidic conditions followed by neutralization, whereas recombinant AAT shows some loss of activity at a wide pH range, which may reflect a different glycosylation pattern. The conditions of the infant-modeled digestion, pH 5 during pepsin treatment, are not ideal for pepsin, which normally possesses full activity at pH 2. AAT has been detected in human infant feces, which supports the notion that it is capable of surviving digestion *in vivo*, particularly during the first three months of the infant's life. This evidence also supports the validity of the *in vitro* digestion system. It is likely that AAT possesses enough resistance to acidic and digestive conditions to allow a significant amount to survive and affect the digestion process.

Recombinant AAT remained functionally intact after being exposed to low pH, *in vitro* digestion, and several types of heat treatment. It is therefore possible that recombinant AAT may be added to infant formula, can tolerate some processing conditions, and remain intact in the gastrointestinal tract of infants. Thus, recombinant AAT may help protect other physiologically active proteins, such as lactoferrin and lysozyme, which also may be added in recombinant forms in the gut of formula-fed infants. In conclusion, addition of recombinant AAT together with other recombinant proteins may enhance their bioactivity and make the formula more similar to human milk.

J. Expression of AAT in transgenic wheat

The plasmid API282 containing the Bx7 promoter, Bx7 signal peptide and AAT gene, Nos terminator and ampicillin resistance gene was used to transform wheat cells, substantially as in the transformation of rice cells. Twenty one transgenic lines were

produced. Expression of AAT was determined to be about 5 to 12 μg per grain of wheat seeds.

EXAMPLE 6

Generation and Characterization of Recombinant Proteins Produced by Transgenic Rice Plants

A. Generation of recombinant antibodies

Recombinant antibodies have been expressed in transgenic plants (for examples, see Peeters *et al.*, 2001; Giddings *et al.*, 2000; Larrick *et al.*, 1998). However, expression and production of recombinant antibodies in the seeds of transgenic plants have certain advantages. The production of high levels of antibodies in grains, for example rice grains, provides distinct advantage that food supplements may be prepared with little or no purification, and other advantages that are illustrated herein the patent application.

In one embodiment, an expression vector is constructed as illustrated in Example 1 that includes codon optimized nucleotide sequences encoding functional components of an antibody. For example, the components can be a heavy chain, a light chain, a linker region or a J chain and a secretory component. The expression vector may also include a promoter, a signal/target/transport sequence or sequences and a terminal sequence or sequences. Preferred promoter, signal/target/transport sequence and terminal sequence are exemplified herein. For example, for expression of each functional component of an antibody in rice seeds, a codon-optimized component gene is operably linked to the rice endosperm specific glutelin (Gt1) promoter, a Gt1 signal peptide and NOS terminator to form a component expression vector.

Each component expression vector is introduced to rice cells and plants to generate antibody component-expressing transgenic rice cells and plants, as exemplified in Example 2. In one embodiment, the expression vectors containing antibody heavy chain, light chain, linker region or J chain, and a secretory component can be introduced individually. The plants expressing each individual component can be crossed to generate plants that express a functional antibody.

In another embodiment, the expression vectors containing functional components of an antibody can be introduced to the plant at the same time, using the transformation methods exemplified in Example 2, such as by co-bombardment. A plant that expresses functional antibody is selected for further propagation.

In another embodiment, the expression vector containing codon optimized nucleotide sequence encoding a single chain antibody is introduced to rice cells and plants to generate antibody expressing transgenic rice cells and plants, as exemplified in Example 2. The nucleotide sequence encoding a single chain antibody can be constructed as conventional in the art, for example Kortt *et al.*, 2001, Maynard and Georgiou, 2000; Humphreys DP and Glover, 2001.

The plant-generated recombinant antibody can be isolated and purified as exemplified in the patent application.

B. Generation of human EGF

The Epidermal Growth Factor (EGF) gene was codon optimized as shown in Figure 20, and synthesized by Operon Technologies (CA, USA) with a SEQ ID NO: 8. The gene was cloned into pAPI145 and pAPI241 respectively, as exemplified in Example 1. The resulting plasmids were named API270 (Figure 21) and API303 (Figure 22), respectively. For expression of EGF in rice seeds, the codon optimized gene was operably linked to the rice endosperm specific glutelin (Gt1) promoter, Gt1 signal peptide and NOS terminator in pAPT303, and to the rice endosperm specific globulin (Glb) promoter, Glb signal peptide and NOS terminator in API270. The transgenic plant expressing EGF was generated, and plant-generated recombinant EGF was detected, as shown in Figure 23 and as exemplified herein.

C. Generation of human IGF

The Insulin-like Growth Factor (IGF) gene was codon optimized as shown in Figure 24, and synthesized by Operon Technologies (CA, USA) with SEQ ID NO: 9. The gene was cloned into pAPI145 and pAPI241 respectively, as exemplified in Example 1. The resulting plasmids were named API271 (Figure 26) and API304 (Figure 25), respectively. For expression of IGF in rice seeds, the codon optimized gene was operably linked to the rice endosperm specific glutelin (Gt1) promoter, Gt1 signal peptide and NOS terminator in pAPI304, and to the rice endosperm specific globulin (Glb) promoter, Glb signal peptide and NOS terminator in API271. The transgenic plant expressing IGF was generated, and plant-generated recombinant IGF was detected as shown in Figure 27 and as exemplified herein.

D. Generation of Other Expression Plasmids

Other expression plasmids for use in transforming plants herein for the production of recombinant polypeptides in transgenic plants were made substantially as previously described. These plasmids are shown in Figure 28, showing API321, containing a Glb promoter, a Gt1 signal peptide, codon-optimized haptocorrin gene, Nos terminator, and an amipicillin resistance gene; Figure 29, showing API320, containing a Gt1 promoter, a Gt1 signal peptide, codon-optimized human haptocorrin gene, Nos terminator, and an amipicillin resistance gene; Figure 30, showing API292, containing a Glb promoter, a Glb signal peptide, kappa-casein gene, Nos terminator, and an amipicillin resistance gene; Figure 31, showing API297, containing a Gt1 promoter, a Gt1 signal peptide, a gene encoding mature kappy-casein polypeptide, Nos terminator, and an amipicillin resistance gene; Figure 32, showing API420, containing a Gt1 promoter, a Gt1 signal peptide, lactohedrin gene, Nos terminator, and a kanamycin resistance gene; Figure 33, showing API418, containing a Gt1 promoter, a Gt1 signal peptide, lactoperoxidase gene minus the sequence encoding the propeptide, Nos terminator, and a kanamycin resistance gene; Figure 34, showing API416, containing a rice Gt1 promoter, a Gt1 signal peptide, codon-optimized lactoperoxidase gene, Nos terminator, and a kanamycin resistance gene; and Figure 35, showing API230, containing a Bx7 promoter, a Gt1 signal peptide, codon-optimized lysozyme gene, Nos terminator, and an amipicillin resistance gene; Figure 36A, showing API254, containing a Glb promoter, a Glb signal peptide, lactoferrin gene, Nos terminator, and an amipicillin resistance gene; Figure 36B, showing API264, containing a Glb promoter, a Glb signal peptide, human lysozyme gene, Nos terminator, and an amipicillin resistance gene; Figure 37, showing API225, containing a GT3 promoter, a Gt1 signal peptide, codon-optimized lysozyme gene, Nos terminator, and an amipicillin resistance gene; and Figure 38, showing API229, containing a RP-6 promoter, a Gt1 signal peptide, codon-optimized lysozyme gene, Nos terminator, and an amipicillin resistance gene.

EXAMPLE 7

Comparison of Promoter Activity in the Expression of Lysozyme in Transgenic Rice

A. Comparison between Gt1 and Glb promoters and signal peptides

In earlier studies, inconsistencies were observed between promoter activity of *Glb* and *Gt1* from transient assay data and the protein accumulation level in transgenic plants bearing the same promoters with signal peptides. These unpublished studies suggested that post-translational regulation was involved in recombinant protein

expression and accumulation in the endosperm. It was unknown whether the storage protein signal peptide played a role in recombinant protein expression level or whether heterologous proteins could be sent to the protein bodies along the sorting pathways of native storage proteins. In order to improve the expression level of recombinant proteins in cereal crop seed, it is important to understand recombinant protein targeting and trafficking in the endosperm expression system. Hence, comparison was made between the rice storage protein promoters and signal peptides from the Glutelin-1 gene ("*Gt1*") and the globulin gene ("*Glb*") showed that both promoters and both signal peptides were capable of effecting expression of lysozyme.

(i). Storage proteins

Rice endosperm contains four main storage proteins: acid-soluble glutelin, alcohol- soluble prolamin, water-soluble albumin and salt-soluble globulin (Juliano BO. Polysaccharides, proteins, and lipids of rice. Am. Assoc. Cereal Chem., St. Paul, MN (1985)). They are targeted into two types of protein bodies in rice endosperm. Prolamin aggregates within the endoplasmic reticulum ("ER") lumen into regularly shaped vacuole called protein body type I. The formation of these protein bodies is dependent on the chaperone BiP80 in the ER. Glutelin is deposited into protein storage vacuoles (PSV) via the Golgi apparatus into irregularly shaped vacuole called type II protein body. The components in the protein body type II and its sorting pathway are not well known. The targeting locations and sorting pathway of globulin and albumin also remain unknown. It appears that once the signal sequence is removed in the ER, the sorting and trafficking depend on the targeting information within the polypeptides and chaperones in the ER. The sorting signals are divided into three categories: sequence-specific vacuole sorting signals (ssVSS), C-terminal vacuole sorting signal (ctVSS), and physical structure vacuole sorting signals (psVSS), as described in Frigerio L. *et al.*, Plant Physiol. 126: 167-175 (2001); Matsuoka K. *et al.*, J. Exp. Botany 50: 165-174 (1999) and Vitale A. & Raikhel N.V., Trends in Plant Science 4: 149-155 (1999).

(ii). Method

Two promoters storage protein genes, *Gt1* and *Glb*, and the corresponding glutelin-1 and globulin signal peptide coding sequences were used to express the human lysozyme protein in developing endosperm. In the three plasmids, pAPI264, pAPI159 and pAPI228, the human lysozyme gene was fused with the nucleotide sequences of the *Glb* promoter and globulin signal peptide coding sequences, the *Gt1* promoter and glutelin signal peptide coding sequences and the combination of the *Glb* promoter with the glutelin (*GT1)* signal peptide coding sequences, respectively (Figure

39A). Lysozyme amounts of T1 seeds were determined for 23 independently transformant lines of pAPI264, 10 lines of pAPI159 and 7 lines of pAPI159. The transgenic lines of pAPI159, which synthesized lysozyme using the *Gt1* promoter and the glutelin signal peptide, produced the enzyme in amounts ranging from 34.25 μg to 297.23 $\mu g \cdot mg^{-1}$ total soluble protein (TSP) with an average of 133.76 $\mu g \cdot mg^{-1}$ TSP. Plants transformed with pAPI264 carrying the *Glb* promoter and the globulin signal peptide yielded between 4.09 and 63.64 $\mu g \cdot mg^{-1}$ TSP lysozyme with an average of 33.96 $\mu g \cdot mg^{-1}$ TSP, while lines of pAPI228, which combined the *Glb* promoter and the glutelin signal peptide, yielded between 8.9 and 203.46 $\mu g \cdot mg^{-1}$ TSP with an average of 87.70 $\mu g \cdot mg^{-1}$ TSP.

The lysozyme expression amounts achieved with the *Gt1* promoter+GT1 signal peptide was 3.94 fold higher than that with the *Glb* promoter+GLB signal peptide, while the expression amounts of lysozyme obtained with the *Glb* promoter+GT1 signal peptide was intermediate but increased 2.58 fold over that produced with the GLB signal peptide (Fig. 39B). Apparently the GT1 signal peptide is more efficient than the GLB signal peptide at lysozyme expression and deposition in rice endosperm. This demonstrates the importance of choosing an optimal signal peptide for the production of recombinant proteins in developing rice grains.

(iii). Chimeric Gene Components

Time course of human lysozyme expression during rice endosperm development. We monitored lysozyme accumulation during endosperm development of transgenic lines 159-1-53-16-1 and 264-92-6. Immature spikelets were harvested at 7, 14, 21, 28, 35, 42 and 49 days after pollination ("DAP"). The lysozyme amounts in the endosperms were measured by the activity assay. Lysozyme accumulation in the seeds of transgenic plant 159-1-53-16-1 began at 7 DAP and peaked at 21 DAP. Thereafter lysozyme content decreased until 35 DAP and then stabilized until seed maturity (Figure 40). Lysozyme accumulation in developing seeds of the transgenic plant 264-92-6 likewise began at 7 DAP, peaked at 28 DAP, after which lysozyme content steadily decreased through seed maturation (Figure 40). These results show that lysozyme accumulation in the two types of transgenic lines during endosperm development follows the same pattern as that of the native globulin and glutelin storage proteins.

(iv). Subcellular localization of human lysozyme in transgenic seeds

In order to determine whether the recombinant lysozyme was targeted to protein bodies in the endosperm, we investigated its subcellular localization by immunofluorescence microscopy. Transgenic plant 264-92-6 synthesizing lysozyme

with the *Glb* promoter and globulin signal peptide and transgenic plant 159-1-53-16-1, producing human lysozyme with the *Gt-1* promoter and the glutelin signal peptide were analyzed. Dual localization with either native glutelin or globulin was used to determine the site of lysozyme deposition.

Synthetic peptides derived from the amino acid sequences of rice glutelin and globulin were used to raise antibodies in rabbits. Antibody specificity was confirmed with Western blots of endosperm proteins. No cross-reaction of glutelin and globulin antibodies with other endosperm proteins was detected with the host TP309 or the transgenic lines 264-92-6 and 159-1-53-16-1. The human lysozyme specific antibody detected the 13 kD of lysozyme protein exclusively in the fractionated and total protein extracts.

Immature seeds from two transgenic lines, 159-1-53-16-1(T4) and 264-92-6(T2) and untransformed control, TP309, were harvested at 14 DAP and fixed and a comparable analysis was conducted. In transgenic line 264-92-6, strong immunofluorescence signals of lysozyme and native proteins were detected with fully overlapping pattern, both when lysozyme and globulin or lysozyme and glutelin were compared (data not shown). Merging the two separately recorded images produces a yellow pseudo color signal. A scan for green and red wavelength emission across the 5 protein bodies along the white line in the not quite perfectly aligned images identifies the co-localization of human lysozyme with globulin. The orange tinge of the protein bodies is due to the stronger emission of red fluorescence than green. The perfect image merger provides a bright yellow color, and the recording of the green and red fluorescence emission along the white line identifies the same 5 protein bodies. These results demonstrated that lysozyme was colocalized in protein bodies with the native storage proteins. The results also demonstrated that the storage proteins globulin and glutelin are localized in the same cell compartment, substantiating the indication that globulin and glutelin are targeted into the same type II protein bodies in rice endosperm. We conclude that lysozyme contains all sorting information for protein body targeting, at least when co-expressed with rice storage proteins.

The localization patterns of lysozyme and native storage proteins in 159-1-53-16-1 are, however, quite different and more complex than those of transgenic line 264-92-6. In transgenic 159-1-53-16-1, lysozyme does not completely colocalize with the native storage proteins. Globulin localized preferentially in smaller, peripheral protein bodies in the younger cells of the cortical region from 14 DAP endosperm, while lysozyme localized preferentially in irregularly shaped protein bodies. However, lysozyme did

colocalize more evenly with globulin in the older cells of the central region from the developing endosperm. Merging the two separately scanned images visualized green fluorescing, lysozyme-rich type II protein bodies and red fluorescing, smaller, globulin-rich protein bodies. Recording of the red and green fluorescence emission along the white scanning line reveals that there is almost twice as much lysozyme in the large type II protein bodies as there is in the small protein bodies, while the globulin signal in the small protein bodies is 2-3 times observed in type II protein bodies. Thus, there appears to be a preferential targeting of the two proteins. In the central region of the endosperm, a more equal co-localization of lysozyme and globulin was observed, especially in the larger type II protein bodies, when judged by the intensity of green and red fluorescence, which provides the yellow color upon merging the two images. This is evident from the merged image scan at the two emission wavelengths. However, there are also small protein bodies containing a dominant portion of globulin or lysozyme in these cells.

Distinct patterns were also found in 159-1-53-16-1 when anti-glutelin antibody was co-incubated with anti-lysozyme antibody in the younger cells of the cortical region from mid-developing endosperm. Like globulin, most of the glutelin localized in the smaller, peripheral protein bodies in younger cells, while lysozyme localized in irregularly shaped protein bodies. Lysozyme partially colocalized with glutelin in the older cells from the center region of mid-developing endosperm. Merging the two images and scanning for fluorescence-intensity at the two wavelengths reveals co-localization of the two proteins in the large and small protein bodies, some being highly enriched in lysozyme and others in glutelin. A comparable distribution is observed in the cells of the central part of the endosperm. The results suggested that lysozyme distorted the native storage protein targeting/sorting when under the control of the *Gt1* promoter/GT signal peptide, producing high lysozyme expression, but not when under the control of *Glb* promoter/ GLB signal peptide with lower lysozyme expression.

To determine if native protein accumulation was affected in the endosperm of the transgenic plants, we analyzed the amounts of glutelin, globulin and lysozyme proteins from two transgenic lines and TP309 by Western blotting. The results showed that glutelin protein was reduced in 159-1-53-16-1, but was increased in 264-92-6 in comparison to TP309. Amounts of globulin protein were reduced in both 264-92-6 and 159-1-53-16-1. This change is particularly significant in the transgenic line 159-1-53-16-1 with its higher lysozyme expression level. The results showed that globulin was more affected than glutelin, no matter which signal peptide was used.

Based on the results, we conclude that lysozyme was targeted to the protein bodies and that the signal peptide played an important role in lysozyme expression. The plants with high expression levels of the recombinant protein showed distorted native protein expression and trafficking.

Thus, the combination of the *Gt1* promoter and *Gt1* signal peptide was more effective than the combination of the *Glb* promoter and *Glb* signal peptide, with the combination of *Glb* promoter and *Gt1* signal peptide having intermediate level of activity. Results showed that the high level expression of recombinant protein distorted the trafficking and sorting of the native storage proteins and affected the native storage protein expression. Results also indicated that mature human lysozyme protein contains a determinant recognized in the plant cell for the protein storage vacuole (PSV) sorting following signal peptide cleavage, and that the lysozyme was sorted to Type II protein bodies.

B. Comparison of seven promoters and Gt1 signal peptide in regulating the expression of lysozyme

Plasmids API159 (Gt1 promoter) (Figure 1), API228 (Glb promoter)(Figure 39), API230 (Figure 35), API229 (RP-6 promoter) (Figure 38), API225 (GT3 promoter), a plasmid carry the Glub-2 promoter, and another plasmid carrying the Club-1 promoter, were compared in their ability to effect the expression of lysozyme in transgenic rice T1 seeds. Results shown in Figure 41 indicate that for expression of lyzoyme, Gt1 was the strongest promoter, followed by Glb, Glub-2, Bx7, Gt3, Glub-1 and Rp6, in order of promoter strength.

EXAMPLE 8

Co-transformation of Heterologous Polypeptide and Reb Gene in Transgenic Rice Plants

A. Enhanced Lysozyme Expression in Transgenic Rice Seed Co-Transformed with Reb

Codon-optimized human lysozyme gene was linked to *Glb* promoter and *Glb* signal peptide to generate plasmid *Glb-Lys* (API264) as shown in Figure 36B, which was used to transform rice with or without *Native-Reb,* as previously described and as described in WO 01/83792. Normal plant phenotypes were obtained among transformants containing *Glb-Lys* alone or with *Native-Reb*. To determine the presence of *Reb* gene and *Glb-Lys* in the transgenic rice genome, one primer designed from

vector sequence and another designed from the *Reb* gene 3' terminator were used to identify these lines. In this case, only the recombinant *Reb* gene could be amplified. PCR analysis confirmed the presence of transgenes in the rice genome. Ten of 11 plants from independent transformation events contained both *Reb* and the lysozyme transgenes. The REB protein of immature seeds from five randomly selected transgenic lines was detected by Western blotting. The expression level of the REB protein in transgenic lines ranged from 25% to 71% higher than that in untransformed TP309. This demonstrated that the transgenic *Reb* gene was active in transgenic plants.

Seeds of confirmed transgenic rice plants were harvested at maturity, and the lysozyme activity was analyzed. As shown in FIG. 19, lysozyme expression in the seeds from 30 independent transformation events containing both the *Native-Reg* and the *Glb-Lys* ranged from 30.57 to 279.61 μg/mg TSP with an average of 125.75 ± 68.65 μg/mg TSP. Seeds of 17 transgenic events containing the *Glb-Lys* gene alone expressed lysozyme in amounts ranging from 7 to 76 μg/mg TSP with an average of 33.95 ± 20.55 μg/mg TSP. No lysozyme activity was detected in untransformed rice seeds. The results showed that the expression level of lysozyme increased an average of 3.7-fold when seeds were transgenic for both the *Reb* gene and *Glb-Lys*. Statistical analysis (t test) showed that the amount of lysozyme in seeds from the plants transgenic for the *Reb* gene and *Glb-Lys* is significantly higher than that in the plants transgenic for *Glb-Lys* alone ($p < 0.001$).

B. Enhanced Human Lysozyme Expression in Transgenic Rice Seed Co-Transformed with Maize Transcriptional Factor, Prolamin-Box Binding Factor (PBF)

Three transcriptional factors were tested; rice endosperm bZIP protein (REB), Opaque2 (O2) and PBF. The transcriptional factors and human lysozyme gene under the control of rice glutelin 1 (Gt1) or globulin (Glb) promoter were co-bombarded into rice callus. Transgenic R_1 grains carrying both genes were obtained. The effect of transcriptional factors on the expression of human lysozyme were monitored. Under the control of Glb promoter, REB increased Lys expression by about 3-fold. REB showed no effect on a stronger promoter, Gt1. Transcription factor increased *Lys* expression, but not significantly. PBF increased *Lys* expression on average 1.5-fold over *Gt1-Lys* alone. The highest *Lys*-expressing lines were selected and advanced to R_2 generation in the greenhouse. As shown in Table 8 below, *Lys* expression level from an R_2 line, 265/159-41-5, was about 190 μg per grain and 9.5 mg/gram of brown rice flour (equivalent to 0.8% grain weight). The level of expression was about 1.5-fold higher

than that of the highest expression line without the transcription factor. In addition, data showed that PBF not only increased the expression of *Lys*, but also increased the expression of native storage proteins such as glutelin and globulin, and the protein related to protein trafficking. It implies that PBF can act on the promoters of multiple genes to increase the expression of those proteins in rice endosperm.

Table 8.

Line Number	R_1* (μg/grain)	R_2* (μg/grain)	R_2 Lysozyme (mg/g brown rice)	Homozygous
285/159-41	150.23	190.00	9.5	homozygous
285/159-43	114.30	155.00	9.0	homozygous
285/159-81	103.58	175.59		heterozygous
285/159-286	152.07	180.00		heterozygous

* The expression data were averaged from 10 seeds in R_1 and from 10 lines in R_2.

EXAMPLE 9

Expression of heterologous polypeptide in transgenic rice under the control of barley aleurone- and embryo-specific promoters

A. Construction of aleurone-specific promoter/GUS fusion plasmids

To obtain *Ltp1* and *Chi26* promoters, PCR forward primers were designed to cover about 900 bp upstream region of each promoter and reverse primers were designed to include the whole 5′ untranslated region. PCR amplification of the 5′ flanking region of *Chi26* from barley genomic DNA (Himalaya) gave a 882 bp PCR product, called PCR-Chi26. Sequence alignments of PCR-Chi26 to Chi26 DNA sequences in Genbank (accession number: L34210) showed deletions/insertions and nucleotide mismatches (FIG. 28A). The PCR-Chi26 promoter with 5′ UTL was transcriptionally fused to the βglucuronidase reporter gene ("GUS") followed by a NOS terminator, resulting in plasmid Chi26/GUS/NOS (FIG. 28B).

Variations were observed in PCR fragments for the *Ltp1* promoter as well. Among the seven varieties of barley, the size of PCR fragments from Harrington and Himalaya was about 500 bp while the expected 957 bp fragment was observed from Klages, Triumph, Golden Promise and Alexis. Both sizes of the PCR fragments were present in I202 (FIG. 29A). PCR fragment amplified from genomic DNA of barley variety Triumph was cloned and called PCR-Ltp1. Sequence alignment of PCR-Ltp1 to *Ltp1* promoter in GenBank (Accession number: X60292) shows only four mismatches (FIG. 29B). The PCR-Ltp1 with 5′ UTL was transcriptionally fused to the β-glucuronidase reporter gene with the NOS terminator, resulting in plasmid Ltp1/GUS/NOS (Fig. 29C).

B. Southern analysis of transgenic plants

Transgenic plants containing each GUS chimeric construct were obtained from rice callus transformed by particle bombardment. A total of 13 transgenic plants were obtained for plasmid Chi26/GUS/NOS and 10 plants for plasmid Ltp1/GUS/NOS. These transgenic plants were confirmed by PCR analysis. To provide physical evidence for the integration of both plasmids into the rice genome, Southern analysis was carried out on representative transgenic rice plants. DNA from four rice plants transgenic for Chi26/GUS/NOS was digested with *Hind*III and *Eco*RI, which should generate a 3056 bp fragment containing the PCR-Chi26 promoter, *gus* gene and the Nos terminator (FIG. 28B). All four plants contained the expected 3056 bp fragment while no signal could be detected in untransformed TP309 plants (FIG. 30A). Similarly, DNA from five rice plants transgenic for Ltp1/GUS/NOS was digested with *Hind*III and *Eco*RI, which should generate a 2935 bp fragment containing PCR-Ltp1 promoter, *gus* gene and the Nos terminator (FIG. 29C). All plants contained the expected 2935 bp fragment, which could not be seen in untransformed TP309 plant (FIG. 30B). In both FIG. 30A and 30B, multiple integration of the GUS chimeric genes into the rice genome was observed. Multiple integration is a common phenomenon in transformation through the particle bombardment (Cao, J. et al., Plant Cell Rep 11: 586-591 (1992); Christou, P. et al., Biotechnology 9: 957-962 (1991); Gordonkamm, W.J. et al., Plant Cell 2: 603-618 (1990).; Tada, Y. et al., EMBO J. 10: 1803-1808 (1991).

C. Tissue-specific expression of GUS in transgenic rice

Transgenic rice tissues were histochemically examined for cell-specific expression of *gus*. Positive GUS expression (blue color) was similar for all transgenic grains. Typical GUS expression derived from transgenic line 217-8-1, transformed with Chi26/GUS/NOS, and line 220-10-1, transformed with Ltp1/GUS/NOS was obserbed. GUS expression under the control of the *Chi26* promoter was localized to the aleurone cells or cells near aleurone layers as evidenced by the blue colortherein (data not shown)). No GUS activity was observed in any other tissues such as leaf, root or stem. To determine if GUS expression was in the aleurone cells, GUS stained grains from transgenic rice containing Chi26/GUS/NOS were sectioned and examined under a dark-field microscope. Under this condition, GUS activity was indicated as red/pink color. In rice, there are one to two aleurone layers (thin aleurone layer) on the ventral side of the endosperm and four to six layers (thick aleurone layer) on the dorsal side which can be anatomically distinguished from other cells [Matsuo, T & Hoshikawa K, Science of the

rice plant. Food and Agriculture Policy Research Center, Tokyo (1993)]. The GUS expression from the *Chi26* promoter was observed only in cells in the thin or thick aleurone layers; GUS expression was not present in the embryo nor in the non-aleurone endosperm cells. This indicates that these anatomically differentiated cells in the thin or thick aleurone layers have the same molecular responsiveness to *Chi26* promoter. No GUS activity was observed in untransformed rice plants.

In contrast to the aleurone cell-specific expression controlled by the PCR-Chi26 promoter, GUS expression under the control of the PCR-Ltp1 promoter was confined to the embryo tissue (data not shown). The whole embryo, including scutelum, of the grains from rice transgenic for Ltp1/GUS/NOS was GUS-stained with blue color. A few aleurone cells adjacent to the embryo cells express a low level of GUS. No other tissues such as leaf, stem or root showed GUS staining in the histochemical analyses of transgenic rice. Previously, limited data from Kalla R. et al., Plant J. 6:849-860 (1994) showed that the *Ltp1* promoter directed GUS expression in vascular tissue of transgenic rice stem. They did not show if there was any GUS activity in rice kernels nor did they show which portion of the *Ltp1* promoter was used to control GUS expression.

Results indicate that in transgenic rice the PCR-Chi26 promoter directs gene expression in rice aleurone cells and the PCR-Ltp1 promoter directs gene expression in embryo cells in transgenic rice. These promoters can be used for controlling the expression of recombinant heterologous polypeptides in aleurone cells.

Expression of heterologous polypeptide in transgenic plants such as rice can increase the commercial food value of rice such as by expressing therein polypeptides of nutritional or health benefits, including those having therapeutic or anti-infective properties or activities, regulate the expression of genes responsible for pest or disease resistance in rice grains to improve rice crop fitness against biotic stresses in the environment, and can substantiate transgenic plants as a suitable expression system for production of any proteins, including industrial enzymes, and other uses of recombinant proteins.. The expression system can also be used for a metabolic engineering project such as the golden rice project [Ye, X.D. et al., Science 287: 303-305 (2000)].

It is interesting to note that the two barley aleurone-specific promoters, *Lpt1* and *Chi26*, are effective in driving expression of heterologous polypeptide in the aleurone and embryo of transgenic rice seeds. It is also of interest that these two barley aleurone-specific promoters show different expression patterns in transgenic rice in contrast to their similar expression in barley. The mechanism underlying the tissue specificity of the two promoters is possibly different. Through transient analysis, Leah et

al. (1994) analyzed the *Chi26* promoter with a series of promoter deletions, mutations and gain of function experiments. Their analyses identified a 33 bp sequence between -179 and -147, so called the E-region, functioning as an enhancer in aleurone tissue and as a silencer in leaf tissue. The sequence of E-region was intact in the PCR-Chi26. The sequence of the E-region is missing from the PCR-Ltp1 promoter though the E-region might be present in *Ltp1* as a trimeric dyad symmetry [Leah, R. et al., Plant J. 6: 579-589 (1994)]. Therefore, the aleurone specific expression of the *Ltp1* gene in barley is likely to be mediated by the promoter element (s) and transcriptional complex(es) which differ from those for *Chi26* genes and this difference is reflected in transgenic rice.

Along with the previous study of the barley *Ltp2* promoter [Kalla, R., (1994)], this expression study of the barley Chi26 promoter using transgenic rice plants indicates that a molecular mechanism controlling gene expression in the aleurone cells is conserved between barley and rice.

D. Plasmid Construction

Standard molecular manipulations and procedures were carried out as described in Ausubel et al. (1987) and Hwang, Y.S. et al., Plant Cell Rep 20: 647-654 (2001).

The 882 bp promoter region with 5′ untranslated leader sequence (UTL) of barley *Chitinase 26* was PCR-amplified from genomic DNA of barley variety, Himalaya, using the primers, Chi26/fw: 5′ AACCCTCTCTGCAGTCACCTCCTGTGAAGT 3′ and Chi26/rv: 5′ CGGAGCGATCTAGATGTGCGAGCCAACAAA 3′. The 957 bp of the *Ltp1* promoter including the 5′ UTL was PCR-amplified from the genomic DNA of barley variety, Triumph, using the primers, Ltp1/fw: 5′ CCGTATAAGCATGCTTGGAACAATCTCCAC 3′ and Ltp1/rv: 5′ GCGCGGGTCTAGATCACTTCTTAATCTGT 3′.

All PCR amplification was carried out using the GeneAmp PCR system (model 2400, Perkin-Elmer). PCR reactions were performed in 50 µl of PCR reaction buffer (50 mM KCl, 1.5 $MgCl_2$ 10 mM Tris-HCl, pH 9.0, and 0.1 %Trition X-100) containing 250 µM dNTP, 1 µg of genomic DNA as template, 0.1 µM each of forward and reverse primers, and 0.3 unit of *Taq* polymerase (Perkin-Elmer). After a two min pre-denaturing step at 94 ° C, the reaction was run for 30 cycles with a 30 s denaturation at 94 ° C, a 30 s annealing at 58 ° C, and a two min extension at 72 ° C. The final extension step was five min at 72 ° C, followed by a 4 ° C soaking step. The PCR-amplified fragments were extracted with phenol:chloroform:isoamyl alcohol. The PCR product containing the *Chi26* promoter and 5′UTL was digested with *Pst*I and *Xba*I and subcloned into the

GUS expression cassette of pBI221 (Clontech, CA) to generate Chi26/GUS/NOS (API217). The Ltp1/GUS/NOS construct (API220) was generated in the same way using a different set of restriction enzymes (*Sph*I and *Xba*I).

E. DNA sequence analysis

Plasmid DNAs containing Ltp1/GUS/NOS (API220, FIG. 28B) or Chi26/GUS/NOS (API217, FIG. 29C) were prepared with the Qiagen miniprep kit (Qiagen Inc. Valencia, CA) and the promoter regions of *Ltp1* and *Chi26* were sequenced by ABI automatic sequencer.

F. Transformation

The transgenic rice plants containing Chi26/GUS/NOS or Ltp1/GUS/NOS were produced as described in Chen, L. et al., Plant Cell Rep 18: 25-31 (1998).

G. Southern blotting

About three grams of green leaves were collected from T_1 transgenic plants and ground in liquid nitrogen into a fine powder. The DNA was isolated according to the procedure of Dellaporta, S.L. et al. (1983) and was purified by phenol/chloroform extraction. About 5 μg of rice genomic DNA was digested with *Hind*III and *Eco*RI and used to prepare blots for Southern analysis. The southern blotting was performed as described in Ausubel et al. (1987). The ECL™ direct nucleic acid labeling and detection system (Amersham, USA) was used for analysis.

H. GUS histochemical staining

Mature T_2 grains derived from all fertile transgenic rice T_1 plants were harvested and soaked in ddH_20 for one day at 4 ° C. These grains were sectioned longitudinally with a sharp razor blade into half-grain portions containing half-embryos. The sections were then stained with X-glucuronide solution for 24 hrs according to Jefferson, R.A. et al., Plant Mol. Biol. Rep. 5: 387-405 (1987).

I. Anatomy and microscopy

The GUS stained half grain was fixed under vacuum overnight in a solution composed of 2.4 % glutaraldehyde, 0.3 % paraformaldehyde, and 0.025 M Pipes buffer. After fixation, the material was rinsed three times in 0.025 M Pipes buffer for 25-30 minutes each and dehydrated using a graduated ethanol series of 10 - 100 % for 25-30

minutes for each step. This was followed by resin infiltration with 3:1, 1:1, 1:3 100 % ethanol: Historesin with each of these steps lasting for 24 hrs. Next the material was immersed in 100 % resin where it sat for 10 weeks. After the infiltration period, the material was transferred to an aluminum weigh boat for embedding in 10 ml of Historesin with 0.667 ml of hardener. Once the plastic had set, the material was cut out, trimmed and mounted onto wooden dowels for sectioning. The material was cut 2-3 µm thick and imaged on an Olympus Vanox AH-2 using darkfield optics.

Example 10

Expression of heterologous polypeptide in transgenic rice under the control of a rice embryo-specific gene promoter

A. Plasmid for Expression in Embryo & Aleurone of Transgenic Grain

Plasmid pAPI224 was made in the following manner. A 246bp PCR fragment containing the first 130bp of the *Empl1* coding region was amplified from rice genomic DNA variety M202 with forward primer CATCCCCATCAGCTCAAGCCGC and reverse primer TGCATACCCTCGGCGAGGTTCTCCTGGGC.

This PCR fragment consisting of 116bp 5' flanking region and 130bp of the Em coding region was cloned into pCR-2.1 a TA cloning vector (Invitrogen). This PCR product was amplified and used to probe to a rice BAC library. A BAC clone, 35 B20, containing Em gene was isolated. 5.5kb of the EM promoter from a BAC clone, 35B20, was cloned into pBluescriptII KS+ using a XhoI digest. A 3.0kb HindIII/NotI fragment of this plasmid was cloned into pCR2.1 a TA cloning vector. To this plasmid a 191bp PCR product (digested with NotI/XbaI) containing the Em leader sequence was added. The 3.0kb Em promoter along with the leader sequence was cloned into pBI221 replacing the 35S promoter using a HindIII/XbaI digest resulting plasmid API224 which contains Em promoter, GUS gene and Nos terminator.

The plasmid API224 was used to transform rice as described. The transgenic seeds were analyzed for GUS activity. It was found that GUS activity was located in embryo and aleurone layer only.

B. Cloning of the rice Em protein gene *Emp1*

The promoter region of the rice *Emp1* gene upstream of the translation start site was cloned into plasmid API224 (FIG. 31) by conventional techniques. The cloning of the rice *Emp1* gene was described in Litts, J.C. et al. , Plant Mol. Bio. 19:335-337

(1992). The *Emp1* gene sequence from Litts, J.C. et al. is reproduced herein as FIG. 53. Plasmid API224 further contained the β-glucuronidase reporter gene and Nos terminator. Rice cells were transformed with API224 as described previously and one transgenic line was studied in greater detail histochemically, as described above.

As shown in FIG. 54, Emp1/GUS expression was observed in transgenic line 257-10-1 in the embryo (panel a), thick layer aleurone (panel b), and thin layer aleurone (panel c). No GUS expression was observed in the seeds of control untransformed rice plant Taipei 309 in any of the embryo (panel d), thick layer aleurone (panel e) or thin layer aleurone (panel f) tissues.

Tissue specificity of GUS expression under the rice *Emp1* promoter was also studied by histochemical analysis, substantially in the same manner as described above. As shown in FIG. 55, GUS expression was observed in transgenic line 257-10-1 in the mature seed (panel a), and in the mature embryo (panel b). No GUS expression was observed in the leaf section (panel c), root section (panel d) or stem cross-section (panel e). No GUS expression was observed in the control untransformed rice seed of Taipei 309, used as control.

EXAMPLE 11

Production of rice extract containing recombinant proteins and its use in food

A. General procedure for production of rice extract

Transgenic rice containing heterologous polypeptides can be converted to rice extracts by either a dry milling or wet milling process. In the dry milling process, transgenic paddy rice seeds containing the heterologous polypeptides, such as recombinant human lysozyme or lactoferrin were dehusked with a dehusker. The rice was grounded into a fine flour though a dry milling process, for example, in one experiment, at speed 3 of a model 91 Kitchen Mill from K-TEC. Phosphate buffered saline ("PBS"), containing 0.135 N NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.7 mM KH2PO4, at pH 7.4, with or without additional NaCl, such as 0.35 N NaCl, was added to the rice flour. In some experiments, approximately 10 ml of extraction buffer was used for each 1 g of flour. In other experiments, the initial flour/buffer ratio varied over a range such as 1 g/40 ml to 1 g/10 ml. The mixture was incubated at room temperature with gentle shaking for 1 hr. In other experiments, the incubation temperature was lower or higher, such as from about 22°C to about 60°C, and the incubation time was longer or shorter, such as from about 10 minutes to about 24 hr. A Thermolyne VariMix platform mixer set at high speed was used to keep the particulates suspended.

In place of PBS, other buffers were used in some experiments, such as ammonium bicarbonate. In one embodiment, 10 liters of 0.5M ammonium bicarbonate was added to 1 kg of rice flour.

The resulting homogenate was clarified either by filtration or centrifugation. For the filtration method, the mixture was allowed to settle for about 30 minutes at room temperature, after which the homogenate was collected and filtered. Filters in three different configurations were purchased from Pall Gemansciences and used. They were: a 3 μm pleated capsule, a 1.2 μm serum capsule and a Suporcap capsule 50 (0.2 μm). For centrifugation, a Beckman J2-HC centrifuge was used and the mixture was centrifuged at 30,000 g at 4°C for about 1 hr. The supernatant was kept and the pellet was discarded.

In one embodiment, the filtrate and supernatant were further processed, for example by ultra-filtration or dialysis or both to remove components such as lipids, sugars and salt.

The filtrate from the above filtration procedure, which is also called the clarified extract, was then concentrated using a spiral wound tangential flow filter operated in a batch recirculation mode. In one embodiment, PES (polyethersulfone) 3000-4000 molecular weight cutoff membranes was used for this step. These final concentrated extracts were held overnight in a cold room.

The concentrated extracts were next dried to a powder by lyophilization. During loading of the lyophilizer trays, the extracts were not subjected to a final 0.2 or 0.45 micron depth filtration to minimize loss of target proteins. The lyophilized material was scraped from the lyophilizer trays and combined into a plastic bag. The dry material was compressed by drawing a vacuum on the bag and then the material was blended and the particle size reduced by hand-kneading it through the plastic.

The lyophilized materials were then suitable for use as an extract directly or in admixture with other food. In one experiment, the lyophilized materials were blended with various ingredients to produce control and test infant-formula. The ingredients were blended using a Hobart mixer (140 quart size) equipped with a paddle agitator. These final blends were packed in 1 kg double Mylar bags and the headspace was filled with nitrogen before sealing.

Table 9 shows the recovery of recombinant human lactoferrin from 105 kg transgenic rice flour during each extraction step. The amount of recombinant human LF present was determined quantitatively as described in Example 4.

Table 9.

	0Lactoferrin	
Stage of process	Lac mass	% of max
Baseline extraction yield	4.0 mg/g flour	
Expected maximum	420 g	100 %
Initial extract	338 g	80
Clarified extract	373 g	89
Concentrated extract	343 g	82
Dried extract	340 g	81

Rice extract can also be produced using a wet milling procedure. Transgenic paddy rice seeds containing recombinant human lysozyme were re-hydrated for a period of 0 to 288 hrs at 30oC. The rehydrated seeds were ground in PBS extraction buffer. The initial seed/buffer ratio varied over a range such as 1 g/40 ml to 1 g/10 ml. Table 10 shows recovery of human lysozyme from rice seeds soaked from 0 to 288 hrs.

Table 10.

Rehydration time (hrs)	Lysozyme (ug/grain)	Recovery (%)
0	87	100
48	69	79
60	79	91
168	60	69
216	56	64
288	58	67

Over 60% human lysozyme was recovered from the wet milling process. The result of the wet milling becomes initial extract which may be stored frozen until use. The processing of initial extract to obtain dried extract was the same as that described for dry milling in this section.

B. Concentration and diafiltration of recombinant lysozyme and control rice extracts

The conditions used in concentration and diafiltration varied depending on volume, speed, cost, etc. These conditions are all routine in the art based on the description herein. The frozen initial extract was thawed in the coldroom (about 2-8°C) for six hours. The thawed material were clarified though a 0.45μm filter and concentrated using a 5000 Nominal Molecular Weight Cutoff membrane of Polyethersultone.

90 ml of the filtrate of control extract was concentrated to 10 ml and additional 10 ml of deionized water was added to the concentrated filtrate. The diluted filtrate was diafiltrated one more time using water. The precipitate started forming at 16 mS and

increased as the ionic strength decreased. 1M ammonium bicarbonate was added to the retentate to add ionic strength. The haze decreased although did not disappear completely. The material was diafiltered multiple times, in one embodiment three times, with water and multiple times, in one embodiment three times, with 0.1 M ammonium bicarbonate. It was concentrated to 9 ml and the membrane was rinsed with 0.1 M ammonium bicarbonate. The concentrate was filtered through several 0.2 μm button filters. In one embodiment, 2.3 ml of the filtrate was lyophilized as is; 2.3 ml of the filtrate was diluted to 12 ml with deionized water and lyophilized, and 2.0 ml of the filtrate was diluted to 25 ml with deionized water and lyophilized. All remained clear.

A total of 89 ml of the filtrate of rHLys extract was concentrated to 10 ml, and additional 10 ml of 0.1 M ammonium bicarbonate was added. The resulting mixture was concentrated back to 10 ml and another 10 ml of 0.1 M ammonium bicarbonate was added. The retentate started to haze up. The material was diafiltered multiple times, in one embodiment three times, with 0.1 M ammonium bicarbonate. It was concentrated to 9 ml and the membrane was rinsed with 0.1 M ammonium bicarbonate. The concentrate was filtered through several 0.45 μm button filters. In one embodiment, 2.0 ml of the filtrate was lyophilized as is; 2.0 ml of the filtrate was diluted to 12 ml with deionized water where a haze formed, and lyophilized, and 2.0 ml of the filtrate was diluted to 12 ml with 0.1 M ammonium bicarbonate which remained clear, and lyophilized.

C. Comparison of trial extraction of recombinant lysozyme rice with PBS and ammonium bicarbonate

The conditions used in concentration and diafiltration varied depending on volume, speed, cost, etc. These conditions are all routine in the art based on the description herein. rHlys rice flour was mixed with extraction buffer at about 100 g/L for about 1 hour using a magnetic stir bar. In one 2 liter beaker, the extraction buffer was PBS, pH7.4 plus 0.35 M NaCl. In another 2 liter beaker, the extraction buffer was 0.5 M ammonium bicarbonate. A 15 cm buchner was pre-coated with about 6g of Cel-pure C300 before adding another 20g of Cel-pure C300. The mixture was filtered at about 3-4 Hg. It was then washed twice with about 100 ml of respective extraction buffer. The extracted filtrate was collected and concentrated with ultra-filtration cartridges: 5K Regenerate Cellulose, 5K PES, and 1K Regenerated Cellulose. The concentrates were lyophilized and analyzed for rHlys contents. The ammonium bicarbonate and PBS, pH7.4 plus 0.35 M NaCl both extracted approximately the same amount of rHlys. There

was little loss of lysozyme units in the permeate with any of the ultrafiltration units that were used.

Other extraction buffer can also be used to extract recombinant proteins expressed in transgenic rice grains, for example Tris buffer, ammonium acetate, depending on applications. For example, for using recombinant human LF for iron supplement, iron may be added to the extraction buffer and the buffer is set at a pH so that the apo-LF can pick up iron during the extraction process. Under this condition, LF can become saturated with iron (holo-LF). In another example, a buffer lacking of iron and a pH resulting in iron release from LF is used to produce apo-LF.

D. Production of rice extracts containing recombinant proteins

The conditions used in concentration and diafiltration varied depending on volume, speed, cost, etc. These conditions are all routine in the art based on the description herein. All equipment was soaked in hot 0.1M NaOH at a starting temperature of about 55°C. Rice flour was added to an about 250-500 gal. stainless steel tank containing 0.5M ammonium bicarbonate at about 95-105 g/L. It was mixed for about 60-80 minutes at about 9°C.

12 plates of 36 inch filter press C300 were pre-coated with about 3-6 kg Cel-pure C300. About 19-26 g/L of Cel-pure was added to the extract and mixed thoroughly. The mixture was pressed at a pressure of about 22 psi at a flow rate of about 82 liters/minute. The filtrate was collected into a 250 gal. stainless steel tank and washed with 0.5M ammonium bicarbonate. The press was blown dry. The process was carried out at about 10°C.

The 300 NMW cut-off membranes (Polysulfone) which had been cleaned and stored with 0.1M NaOH after control run was rinsed thoroughly with deionized water. The extract was concentrated and bumped to a 100 gal stainless steel tank. The membrane and the concentration tank were flushed with 0.1M ammonium bicarbonate to recover all of the products. The product were covered with plastic and left in the 100 gal tank overnight at room temperature. The concentrate was filtered through spiral wound 1 μm filter and into 5 gal poly container. The concentrate was lyophilized. About 81% of lactoferrin and about 58% of lysozyme was recovered from transgenic rice grains, respectively.

E. Blending of rice extract containing recombinant proteins into infant formula

The three types of lyophilized dry extract that contains rice proteins (control) or rice proteins with lysozyme or lactoferrin were combined with standard infant formula. The blending was done such the final infant formula contained about 1 gram lactoferrin and 0.1 gram lysozyme per liter of infant formula. The ingredients were blended using a Hobart mixer (140 quart size) equipped with a paddle agitator. These final blends were packed in 1 kg double Mylar bags and the headspace was filled with nitrogen before sealing.

Samples of infant formula containing human lysozyme and lactoferrin were quantified using procedures described in Example 3 and 4.

Table 11 shows human lysozyme and lactoferrin in infant formula.

Table 11.

Infant Formula	Lactoferrin (mg/ ml)	Lysozyme (mg/ml)
With control rice extract	0.0	0.0
With transgenic rice extract	1.03	0.13

Using extract as a delivery method of recombinant protein has clear advantages over the purified form or in the whole grain. The conventional approach, such as in the whole grain form, has limitations such as protein stability during high temperature and pressure processing. Furthermore, the purification approach is expensive. Therefore the extract approach 1) maintains a low cost compared to purification approach; 2) requires much smaller volume, for example about 1-10% of whole grain weight; 3) increases the concentration of recombinant protein from about 0.05-0.5% in whole grain form to about 10 to 20 % in the extract form. Some extract form even reaches 40% depending on the expression level of recombinant protein. Therefore, the extract approach will allow broader application of the recombinant proteins compared to the whole grain approach. In addition, the extract approach removes starch granule, which requires high gelling temperature, for example about 75°C. Consequently, the extract approach provides more flexibility in processing the rice grain and the recombinant proteins into food and diet, and the alike, without worrying about using high temperature to denature starch granule. The undenatured starch granule cannot be digested by human gut without gelatinization by for example high temperature.

EXAMPLE 12

Plant Transcription Factors

A. Cloning Of The Reb Gene From A Rice BAC Library

Reb was cloned from a rice BAC library. The Reb gene including the introns, promoter and 3'- UTR region is 6,227 bp long, comprises 6 exons and 5 introns and is flanked by a 1.2 kb 5' promoter and a 1.2 kb 3'-terminator region. The function of the Reb gene was explored using effector constructs containing the Reb gene together with the native Reb promoter and fusion genes linking Reb to the rice actin (Act) or globulin (Glb) gene promoters. (See Fig. 6A-B.)

PCR primers were designed based on the Reb gene sequence provided in Nakase, 1997 and used to screen a rice bacterial artificial chromosome (BAC) library [Yang, 1997] using a screening strategy for tri-dimensional DNA pools of the BAC library as described by Xu, 1998.

PCR was carried out with 100 ng pooled BAC DNA, 10 mM Tris (pH 8.3), 1.5 mM $MgCl_2$, 50 mM KCl, 0.5 μM dNTP and employed a program of denaturing at 94 °C for 5 min followed by 30 cycles of 94 °C for 45 seconds, 60 °C for 45 seconds and 72 °C for 1 min, using a forward primer (5'-CTGATATGTGCCCATGTTCCAAAC-3') and a reverse primer (5'-CCTTGCTGAATGCAGATGTTTCAC -3'). The plasmid DNA of a positive BAC clone was prepared as described [Yang, 1997], the BAC DNA digested with *Hind*III and the presence of the Reb gene confirmed by Southern analysis [Sambrook J., 1989].

The Reb gene was retrieved from the BAC by subcloning two fragments into the pBluescript KS+ vector (Stratagene, CA). First, the promoter and partial coding region was obtained as a *Kpn*I-*Hind*III fragment, followed by a second step where a *Hind*III fragment containing the remaining coding region and the 3' terminator region was obtained by shut-gun cloning. The two fragments were ligated at the internal *Hind*III site generating an intact Reb gene and the complete Reb gene was generated by ligating a 1,775 bp fragment containing the promoter and the 5' coding region, to a 4,452 bp fragment containing the 3' coding and terminator region.

The Reb DNA was sequenced with an automatic DNA sequencer (ABI 371) which revealed 5 introns, 6 exons, 1.16 kb of the 5' promoter sequence and 1.2 kb of the 3' region totaling 6,227 bp. (Fig. 2A-I). A comparison of the open reading frame of the isolated Reb gene with the Reb cDNA gene found at GenBank Accession No. ABO21737 revealed 99.97% DNA sequence similarity and 99.99% amino acid similarity

resulting from two amino acid changes: Ile_{165}→Asn and Glu_{215}→Lys. These differences are likely to be due to polymorphisms among rice varieties.

B. Opaque2 (O2) and Prolamin Box Factor (PBF)

The nucleic acid sequence for the Opaque2 transcription factor found for example at GenBank Accession number X15544 (opaque2 gene), M29411 (opaque2 cDNA) was cloned into an expression vector under the control of the rice glutelin-1 (Gt-1) promoter (Fig. 11C, maize).

The rice PBF and maize PBF coding sequences found for example at GenBank Accession numbers D11385 (rice cDNA) and ZMU82230 (maize cDNA) were also cloned into an expression vector under the control of the rice glutelin-1 (Gt-1) promoter.

The DNA sequence of the rice (Oryza sativa) globulin promoter, ("Glb") with putative binding sites for the O2 transcription factor and the prolamin box and the DNA sequence of the wheat Bx7 promoter with putative binding sites for the O2 transcription factor and the prolamin box are shown in Figs. 9 and 10, respectively.

Promoters were digested to produce the appropriate cohesive ends and cloned into compatible sites in a reporter construct. In one example, the reporter construct is comprised of the rice globulin (Glb) or wheat Bx7 promoter translationally fused with GUS, with the resulting constructs designated, Glb/GUS and Bx7/GUS (Fig. 12A).

The effect of plant transcription factors on the gene expression was evaluated by co-transformation with a heterologous nucleic acid construct effective to express the transcription factor, *e.g.* O2, PBF, BPBF or Reb.

EXAMPLE 13

Transient Expression Assays with the Reb Transcription Factor

A. Plasmid Construction

Plasmids were constructed using standard molecular biological techniques as described in Ausubel *et al.*, 1987. Plasmids API212 (Glb-GUS) carrying the rice globulin gene promoter fused to the GUS reporter gene and API142 (Gt1-GUS) containing the rice glutelin 1 gene promoter fused to the GUS reporter gene were used for transient expression assays. The globulin promoter (GenBank Accession number X63990) and glutelin 1 promoter (GenBank Accession number Y00687) were obtained from *M2O2* (*Oryza sativa Joponica* subsp.) by amplification with PCR. The plasmid containing the Reb gene under the control of its native promoter was designated pAPI267 (Native-Reb). The plasmid designated Glb-Reb (pAPI266) was prepared by cloning the Reb coding region (*NruI*/*SacI* fragment) into the Glb-GUS plasmid after removal of the GUS

gene by digestion with *SmaI/SacI*, thus replacing the GUS gene with the Reb gene. Using the same strategy, the plasmid designated Actin-Reb (pAPI277) was made by replacing the GUS gene of plasmid Act1-D-GUS (McElroy, 1990) with the complete Reb gene (*NruI/SacI* fragment). The Act1-D promoter was kindly provided by Professor Ray Wu, Cornell University (McElroy, 1990).

B. Transient Assay using Rice Endosperm

Rice spikelets with immature endosperm (7-9 days after pollination, dap) of M2O2 (*Oryza sativa Japonica subsp.*) were collected from plants grown in the greenhouse at 30°C. The spikelets were sterilized with 70 % ethanol for 10 min. After evaporation of residual ethanol, the endosperm was dissected and 10 immature endosperms placed on a filter paper in a Petri dish containing AA medium [Chen *et al.*, 1998] supplemented with 20mM ammonium nitrate.

Fifty µl of gold particles (60mg·ml^{-1} at 1:1 ratio of 1.0 and 1,5-3.0 µm diameter gold particle) were coated with 5 µg DNA consisting of a mixture of the reporter gene, the effector gene, and the internal control gene, typically at a molar ratio of 1:1:1. DNA coating was accomplished as described in the instruction manual of the Biolistic PDS system (Bio-RAD, Hercules, CA, USA). pAHC18 containing the luciferase gene driven by an ubiquitin promoter [Christensen *et al.*, 1996] was used as an internal control. In tests without the effector gene, pAHC18 was replaced by pBluescript DNA. Particle bombardment was carried out with a biolistic Helium gun device at 1100 psi (Biolistic PDS 1000/He system, Bio-Rad, Hercules, CA, USA). After bombardment, the immature endosperms were incubated at 25°C for 24h in 5 ml of AA medium supplemented with 20 mM ammonium nitrate, 50µg·ml^{-1} cefotaxin and 50µg·ml^{-1} timentin (Sigma, Louis, MO, USA) to prevent bacterial growth. The endosperms were then harvested and ground with 55 µl extraction buffer (0.1 M potassium phosphate, pH 8.0, 1 mM EDTA, 10 mM DDT, 5 % glycerol, 0.2mM leupeptin and 0.2 µM phenylethylsulfonyl fluoride (PMSF). The extract was centrifuged at 25,000 g for 5 min at 4°C. From the supernatant, a 20 µl aliquot was added to 180 µl of luciferase assay buffer (0.25 M Tricine, pH 7.8, 150 mM magnesium chloride, 10 mM ATP, 1 mM DDT and 100 µg ml^{-1} BSA). Another 20 µl aliquot was added to 200 µl of GUS assay buffer (Tropix, Bedford, MA, USA). Luciferase activity was measured after incubation at 25°C for 20 min and β-glucuronidase activity after 1h at 37°C for 1h with a Monolight 2010 chemi-illuminometer according to the manufacturer's instructions (Analytical Luminescence Lab., Monolight, San Diego, CA, USA). β-glucuronidase activity was normalized to the luciferase activity

and expressed relative to the activity of the Glb-GUS. In general, the data presented reflect an average of at least six assays of two independent experiments.

C. Transcriptional Activation with Reb

The function of Reb gene was analyzed by the transient assay with rice immature endosperm. Transient expression studies were carried out to evaluate the effect of Reb on expression of GUS (β-glucuronidase), under the control of the Glb promoter.

The results shown in Fig. 7B indicate that GUS expression was increased when Reb was expressed under the control of the globulin promoter, the actin promoter or the native Reb promoter suggesting that Reb is effective to an activate expression mediated by Glb promoter.

Putative Reb binding sites in the globulin (Glb) promoter were identified as shown in Fig. 4.

In order to determine, if binding of the Reb protein to the motif (GCCACGT(A/C)AG) in the globulin (Glb) gene promoter activates transcription of this promoter, plasmids containing fusions of the Reb coding region with the Glb promoter and the rice actin (Act) gene promoter were prepared (Fig. 7A). These as well as the expression plasmid containing the native Reb gene (pAPI266) were co-bombarded into the rice endosperm with a plasmid containing the GUS reporter gene driven by the Glb gene promoter and an internal control plasmid containing the luciferase gene driven by the ubiquitin promoter.

Setting the level of GUS expression by the globulin gene promoter as 1, the co-delivery of the plasmids containing the Reb gene increased GUS expression irrespective of whether the gene was driven by its own promoter or the Glb promoter or the Actin promoter (Fig. 7B). The increases were 2.43, 2.01 and 1.98 fold, respectively. The activation of GUS expression was abolished when a promoter-less Reb construct was co-bombarded with Glb-GUS (Fig. 7B). These results suggested that the Reb protein functions as a transcriptional activator of the Glb promoter.

D. Identification of the Upstream Activation Sequence (UAS) for Reb

Using a band-shift assay, Nakase *et al.*, 1997 have shown that Reb binds to two motifs, GCCACGTAAG or GCCACGTCAG. An analysis of the Glb promoter sequence revealed two copies of GCCACGTAAG and one copy of GCCACGTCAG clustered in the sequence region -700 bp distal to the TATA box of the promoter (Fig. 4).

In order to determine whether the binding motifs for Reb signify an upstream activation sequence (UAS), 200 bp of the Glb promoter, which contains the three motifs located at positions -642 to -842 distal to the TATA box were deleted from the Glb promoter (Fig. 8A). The deletion was demonstrated to have no effect on the expression of GUS as both the Glb-GUS and the GlbΔUAS-GUS (UAS deleted) constructs showed the same level of background GUS expression (Fig. 8B).

When Native-Reb was co-bombarded with GlbΔUAS-GUS, the transcriptional activation by Reb which was evident when Native-Reb was co-bombarded with Glb-GUS was lost (Fig. 8B). These data indicate the transcriptional activation of Glb-GUS by Reb occurs through this 200bp fragment containing Reb binding motifs.

A scan of the rice glutelin1 (Gt1) promoter sequence did not reveal the presence of Reb binding motifs. Accordingly, Gt1 was selected as a candidate for the introduction of the UAS from the Glb promoter in order to test for gain of the Reb response function. Heterologous nucleic acid constructs were prepared containing the native Gt1 promoter linked to the GUS gene (Gt1-GUS), and a Gt1 promoter modified to contain a 98 bp Reb UAS fragment containing 3 copies of GCCACGT(C/A)AG (amplified from the Glb promoter) was inserted at position -630 bp distal to the TATA box of the Gt1 promoter in order to generate Gt1+UAS-GUS (Fig. 9A).

When Gt1-GUS was tested by co-bombardment of developing endosperm with the native Reb gene and Gt1-GUS, the results showed that Reb does not activate the Gt1 promoter (Fig. 9B). Gt1+UAS-GUS was tested for GUS expression and it was shown that addition of the Reb UAS to the Gt1 promoter did not increase its capacity for GUS expression significantly (Fig. 9B). However, when Gt1+UAS-GUS was tested by co-bombardment of developing endosperm with the native Reb gene a 2.5 fold increase in GUS activity was obtained (Fig. 9B).

The Reb protein was previously described as a transcription factor. The results described herein show that (1) Reb is a transcriptional activator, as evidenced by a 2.0 to 2.5-fold increase in GUS activity when Reb effector constructs were co-transferred with the reporter *uid* A gene encoding GUS under the control of the Glb promoter into immature rice endosperm cells; (2) Reb specifically activates the Glb promoter but not gluletin gene family promoters; (3) Reb interacts with an approximately 100 bp upstream activation sequence (UAS) containing the motifs GCCACGTCAG and GCCACGTAAG (GCCACGT(A/C)AG) of the Glb promoter, as confirmed by loss-of-function and gain-of-function experiments. The loss of activation function, when the 200bp fragment containing the Reb UAS is removed from the Glb gene promoter, and the gain of this

function, when the 98bp fragment with Reb UAS is added to the Gt1 promoter, establishes the 98bp fragment as an upstream activation sequence (UAS).

EXAMPLE 14

Transient Expression Assays with the O2 and PBF Transcription Factors

A. Plasmid Construction

Plasmids were constructed using standard molecular biological techniques as described in Ausubel *et al.*, 1987. A 693 bp segment of the *Gt1* promoter sequence was removed from pGt1 v3.0 SDM, a *Gt1* expression vector, and used to replace the *CaMV 35S* promoter in pBI221 through *Hind*III and *Sma*I sites, resulting in Gt1/GUS/NOS. The promoter regions from the rice glutelin genes *Gt3*, *GluB-1* and *GluB-2*; the rice prolamin genes *RP6* and *PG5a*; the rice globulin gene *Glb*; and the wheat glutelin gene, *Bx7;* were PCR amplified from M2O2 genomic DNA or the wheat variety, Anza, for *Bx7*.

The PCR amplifications were carried out using the GeneAmp PCR system (model 2400, Perkin-Elmer) operated according to the manufacturers instructions. Basic cycling conditions were 30 cycles, after a 2 minute pre-denaturing step at 95 °C, with a 30 second denaturing step at 95 °C, a 30 second annealing step at specific temperature, and a 2 minute extension step at 72 °C. The final extension step was 5 minute at 72 °C, followed by 4 °C soaking step. Reaction components per 50 µl volume, were 1 µg of genomic DNA or 1 ng of plasmid DNA, 2.5 µl of 5 µM primer mixture, 5 µl of 10 mM dNTP, 2.5 units of *Taq* polymerase (Perkin-Elmer), 5 µl of 10X PCR buffer (Perkin-Elmer). The concentration of $MgCl_2$ was 1.5 mM, for all the promoters with the exception of the *Bx7* promoter for which 2.5 mM $MgCl_2$ was used. All the PCR primers and amplified fragment sizes are presented in Table 12.

Each of the PCR amplified promoter sequences was cloned into the GUS cassette of pBI221 through *Pst*I and *Xba*I sites to give Gt3/GUS/NOS, GluB-1/GUS/NOS, GluB-2/GUS/NOS, RP6/GUS/NOS, PG5a/GUS/NOS, Glb/GUS/NOS, Bx7/GUS/NOS, for the *Gt3*, *GluB-1*, *GluB-2*, *PG5a*, *RP6*, *Glb* and *Bx7* promoters, respectively.

Table 12 Primer sequence used to amplify promoter fragment

Primers	Primer sequences	PCR amplified fragment (bp)	Annealing temperature
Gt3/fw Gt3/rv	GTTAGTcTGCAgTGTAAGTGTAGCTTC ATGGTTGtCtaGaTTTTGTGGGACTGAAC	856	58°C
GluB-1/fw2 GluB-1/rv2	ACAGACAGcTGcAGAGATATGGATTTTC TAAG GGAACTCtCtAgAGCTATTTGTACTTGCT TATG	1319	62°C
GluB-2/fw GluB-2/rv	TCCGAGctgcAGTAATGGATACCTAGT GTAGTTtCtAgAGCTATTAGCAGTTGC	1028	58°C
PG5a/fw2 PG5a/rv2	CGGTGcTGcAGATGGGTTGGGAACCCT ATGATCTagATTGCTCTGGGACATAGAT	874	58°C
RP6/fw RP6/rv	AATTCCTgCagCATCGGCTTAGGTGTA TGATCTagATTGTTGTTGGATTCTACT	684	58°C
Osglb/fw2 OSglb/rv	GGCGCCTGcAGGGAGGAGAGGGGAGA GAT ACCTTGCTctagATTGATGATCAATCAGA	997	58°C
Bx7/fw2 Bx7/rv	CGTCGTCTcTGcAGGCCAGGGAAAGAC AATG CGCTTAtCtAgaTCAGTGAACTGTCAGTG	993	62°C

All transcription factor ("effector") plasmids where the coding sequences were placed under the control of the CaMV 35S promoter were generated by subcloning cDNA fragments of either O2, o2676 or PBF downstream of the CaMV 35S promoter and the *Adh1* intron and upstream of the *nos* 3' end in pMF6. For O2, the *Bgl*II O2Δl cassette described by Schmidt *et al.*, 1992 (which removes the start codons for the three small ORFs present in the 5' leader sequence, thus promoting increased O2 expression), was inserted into the *BamH*I site of pMF6. The o2676 effector was generated as described for O2 except that an internal restriction fragment containing the o2676 point mutation (Aukerman, 1991) was substituted for the corresponding restriction fragment in *Bgl*II O2Δl cassette. For PBF, a *BamH*I-*Xho*I fragment containing the entire PBF cDNA (VicenteCarbajosa, 1997) was subcloned into the same sites of pMF6. O2Δ1 and PBF were also expressed under the control of the maize *UBI1* promoter and first intron by subcloning the respective cDNA clones into the BamHI site of the pAHC17 plasmid (Christensen *et al.*, 1996).

Figures 11A-C are a schematic depiction of plasmids containing the (A) barley prolamin box binding factor protein (BPBF), (B) maize prolamin box binding factor protein (PBF) and (C) the maize opaque2 binding protein (O2) transcription factor

coding sequences under the control of the rice endosperm-specific glutelin promoter (Gt-1).

The construction of antisense plasmids for the O2 and PBF DNA binding domains was carried out using PCR primers designed to amplify the highly conserved region of DNA binding domains of O2 and PBF. The PCR primer sets for O2 were (MO2/fw: 5'-TTCTGGGATCCAAGATGCCTACCGAGG-3') and MO2/rv: 5'-GGGGTCGGATCCGAGATGGGCATGGAC-3', and for PBF (PBF/fw: 5'-AGTGGGGATCCTAAGCCGAGGCCGCAAC-3' and PBF/rv: 5'-GCTAGGGGATCCTGGTGCATAGGTAGCA-3', resulting in amplification of 333 bp and 278 bp amplification fragments, respectively when *Ubi*:*O2* and *Ubi*:PBF were used as the template.

The PCR reactions were performed in 50 µl of 1X PCR reaction buffer (50 mM KCl, 10 mM Tris-HCl, pH 9.0, and 0.1 % Triton X-100) containing 250 µM dNTP, 1 ng of template, 0.1 µM of each of forward and reverse primers, and 0.3 unit of *Taq* polymerase (Perkin-Elmer). The GeneAmp PCR system (model 2400, Perkin-Elmer) was programmed for an initial denaturing temperature of 94 °C for 4 min, a 30 sec denaturing temperature of 94 °C, an annealing temperature of 58 °C for 30 sec, and an extension temperature of 72 °C for 2 min. The reaction was carried out for 30 cycles. An additional extension at 72 °C followed for 5 min was allowed to proceed after completion of the final cycle. The PCR products were purified by phenol:chloroform:isoamyl alcohol extraction procedure and precipitated with 100 % ethanol. After resuspending in 50 µl of dH20, the amplified products were digested with *BamH*I. The *BamH*I flanked PCR products were used to replace the luciferase coding region of pAHC18 [Christensen *et al.*, 1996].

Insertion of the PCR-amplified DNA binding domain of *O2* and PBF in antisense orientation into the expression cassette containing the ubiquitin promoter was completed by PCR using the primer sets MO2/rv and NOS/rv; 5'-CGGCAACAGGATTCAATCT-3', PBF/rv and NOS/rv. The PCR was performed under the conditions described above, with the exception that the annealing temperature was changed to 53 °C.

B. Transient Assay using Rice Endosperm

DNA coated gold particles were prepared by mixing 50 µl of gold suspension (60 mg/ml), 50 µl of $CaCl_2$, 2.5 M and 20 µl of spermidine, 0.1 M. In all cases, 5 µg of the GUS chimeric construct and 5 µg of pAHC18 (containing the luciferase gene under the

control of the ubiquitin promoter), were used. For co-transfection with effector plasmids, effector plasmids were additionally added in the amount indicated. The total amount of plasmids used for coating gold particles remained constant by adding the pBluescriptII KS (+) or the PMF plasmid (containing the *CaMV 35S* promoter driving an expression cassette which lacks a coding region). After vortexing for 1 min, the gold particles were washed with 100 % ethanol twice and finally resuspended in 50 µl of 100 % ethanol. 10 µl of gold particle suspension was loaded into a macrocarrier for bombardment.

Rice immature seeds at 7 to 9 DAP (days after pollination) were harvested and sterilized with husk by incubating 10 minutes in 70 % ethanol and followed by spraying with 100 % ethanol. After the ethanol evaporated completely, transient assay incubation buffer (TAIB: complete AA medium [Thompson, 1986] supplemented with NH_4NO_3 1.4 g/L, 100µg/ml of cefotaxime and 100 µg/ml of timentin) was added to prevent seeds from drying out. A portion of each seed grain which contains the embryo (about one fifth) was cut off using a sharp blade and the immature endosperm was squeezed out. About 10 rice immature endosperms were placed in the center of a 3 MM Whatman filter paper, prewetted with TAIB. Particle bombardment was carried out using the biolistic helium gun device (Dupont PSD-1000), as described Hwang *et al.*, 1998. After bombardment, the bombarded immature endosperms were incubated in 5 ml of TAIB in the dark at 25 °C for 1 day.

After incubation, immature endosperms were transferred to a conical Eppendorf tube and homogenized in 55 µl of extraction buffer (KH_2PO_4, 0.1 M, EDTA, 1 mM and β-mercaptoethanol, 7 mM) by a disposable plastic pestle. After spinning down the cell debris by centrifuging at 15,000 rpm for 15 min, 20 µl of supernatant was used in an assay for GUS or luciferase enzyme activity [deWet, 1987; Bronstein, 1994].

C. Transcriptional Activation with O2 and PBF

Transient expression assays were carried out to evaluate the effector activity of the O2 and PBF transcription factors on GUS expression.

Rice immature endosperms were isolated from caryopses at 7-9 DAP (DAP:Days After Pollination). The amount of each effector plasmid used is indicated in the individual examples, below. The total amount of plasmid for coating the gold particles remained constant by adding pBluescriptII KS(+). To normalize transfection efficiency, in every experiment, a luciferase gene under the control of the maize ubiquitin promoter was co-transfected (with 5 µg of pAHC18 (Ubi/LUC/NOS) used as an internal control for all experiments). Following measurements of GUS and LUX activity,

the GUS expression level was normalized by dividing by the LUX activity from luciferase to obtain the GUS/LUX ratio. Therefore, the GUS/LUX ratio quantitatively indicates the transcriptional activity from the promoter of the heterologous reporter gene.

Developing rice endosperms at 7-9 DAP were biolistically bombarded with various heterologous nucleic acid constructs to examine the ability of the maize trans acting factor genes encoding opaque 2 (O2) and prolamin box binding factor (PBF) to effect expression of heterologous nucleic acid constructs under the control of the promoter regions from the rice glutelin genes Gt3, GluB-1 and GluB-2; the rice prolamin genes RP6 and PG5a; the rice globulin gene Glb; and the wheat glutelin gene, Bx7. O2 and PBF were expressed under the control of the Ubi promoter in the Ubi:O2 and Ubi:PBF constructs, respectively.

As shown in Figure 13B, co-transfection of the *Gt1* reporter plasmid (Gt1/GUS/NOS) with *35S:O2* and *35S*:PBF resulted in approximately a 5 and 3-fold in transcription activity, respectively, relative to the activity from the *Gt1* reporter plasmid alone (Figure 13B).

When immature endosperms were co-bombarded with an equimolar mixture of the *35S:O2* and *35S*:PBF constructs, GUS expression from Gt1/GUS/NOS increased up to 6-fold (Fig. 14B). Transient assays using effector plasmids expressing mutant forms of the O2 and PBF transcription factors demonstrated that the observed activation is a consequence of direct interaction between O2 and PBF and the *Gt1* promoter. The relative GUS/LUX ratio from the *Gt1* reporter plasmid was not affected by co-bombardment with effector plasmids expressing defective forms of the O2 and PBF proteins (Fig. 14B) indicating that the capability of O2 and PBF to bind their specific target sites is critical for transactivation of the *Gt1* promoter.

Figure 13B illustrates transactivation of the *Gt1* promoter by various amounts of O2 and PBF effector plasmids in rice immature endosperms. As shown in Fig 13B, *Ubi:O2* and *Ubi*:PBF at the amount of 1 µg were able to transactivate the *Gt1* promoter, approximately 4 and 3 fold, respectively. This transactivation effect increased along with the increase in the amount of effector up to approximately 5 µg. The increase in promoter activity by O2 and PBF together was shown to be additive, independent of the amount of plasmid combined (Fig. 13B).

As described above, in addition to the promoter of rice glutelin gene (*Gt1*), several other promoters of genes encoding different kinds of seed storage proteins like rice globulin, rice prolamin and wheat glutenin were tested for their responsiveness to O2 and PBF (Fig. 12A). The transcription activity from the promoter of rice globulin, *Glb*

was increased by about 3 and 4 fold, in the presence of O2 and PBF, respectively (Fig. 12B). *Bx7*, the glutelin promoter from wheat, was shown to be transactivated in rice immature endosperms by co-bombardment with O2 and PBF up to about 1.8 fold, respectively (Fig. 12B). The O2 and PBF effectors were also shown to transactivate the rice prolamin gene promoters, *RP6* about 5 and 3.5 fold, and *PG5a*, about 2 and 1.5 fold, respectively (Fig. 12B).

The effect of O2 and PBF on the promoters from different kinds of rice storage genes including *Gt1*, *Gt3*, *GluB-1* and *GluB-2*; the rice prolamin genes *RP6* and *PG5a*; the rice globulin gene *Glb* and the wheat glutelin gene, *Bx7* was tested. Table 2 shows the responsiveness of the promoters to O2 and PBF, with an additive increase in transactivation observed by co-transfection with both effector plasmids. Of the promoters tested, the *Gt1* promoter was the most responsive to both of O2 and PBF and co-transfection with both plasmids gave about a 15 fold increase in GUS activity. The additive increase in promoter activity by co-bombardment of both effector plasmids was observed in all the promoters of storage protein genes that were tested. The rice actin promoter was much less responsiveness to O2 and PBF effectors and the promoter activity of the *CaMV 35S* promoter was not affected by co-transfection with O2 and PBF effector plasmids (Table 13).

Table 13 GUS expression in rice immature endosperms in response to O2 and PBF.

0Effector	O2[1]	PBF[1]	O2/PBF[1]
Gt1	10.8(2.14)	4.57(l.55)	14.6(1.08)
Glb	3.22(0.72)	4.42(0.85)	8.55(2.27)
RP6	5.22(l.45)	3.36(0.71)	8.20(1.18)
Bx7	1.78(0.13)	1.84(0.38)	3.81(l.03)
PG5a	2.13(0.42)	1.59(0.18)	3.14(0.60)
Actin	1.59(0.25)	1.14(0.21)	1.64(0.48)
CaMV 35S	1.18(0.04)	1.06(0.18)	1.34(0.12)

[1] fold activation was calculated by normalizing the GUS/LUX ratio from rice endosperms co-bombarded with each effector and a reporter construct to the GUS/LUX ratio of that effector alone. For example, for GT1, all results are given relative to the GUS/LUX ratio of Gt1/GUS/NOS construct.
S.D. indicates the standard deviation of a mean value for at least five independent particle bombardments.

[1] fold activation was calculated by normalizing the GUS/LUX ratio from rice endosperms co-bombarded with each effector and a reporter construct to the GUS/LUX ratio of that effector alone. For example, for GT1, all results are given relative to the GUS/LUX ratio of Gt1/GUS/NOS construct.
S.D. indicates the standard deviation of a mean value for at least five independent particle bombardments.

[1] Results are presented as "fold activation(SD)", calculated by normalizing the GUS/LUX ratio from rice endosperms co-bombarded with each effector and a reporter construct to the GUS/LUX ratio of that effector alone. For example, for GT1, all results are given relative to the GUS/LUX ratio of Gt1/GUS/NOS construct.
S.D. indicates the standard deviation of a mean value for at least five independent particle bombardments.

EXAMPLE 15

Generation of Transgenic Plants Which Express Plant Transcription Factors

Heterologous nucleic acid constructs were prepared for generation of stable transgenic plant lines. Figs. 15A and B present a schematic depiction of exemplary plasmids for use in generating such stable transgenic plants. The plasmids contain a heterologous protein coding sequence for lactoferrin and lysozyme under the control of the rice endosperm-specific globulin promoter (Glb), as shown in Fig. 15A and Fig. 15B, respectively.

A. Vector Constructs

A gene encoding the mature polypeptide of human lysozyme (EC 3.2.1.17) with a G+C content of 68.4 % was synthesized (Operon, Alameda, CA, USA) based on the sequence of GenBank Accession number J03801. The DNA was digested with *DraI/XhoI*, and ligated into the *NaeI/XhoI* sites of the expression cassette in plasmid API241 which contains the rice globulin gene promoter and signal peptide (GenBank Accession Number X63990). The resulting plasmid was named pAPI264.

B. Development of Stable Transformants and Transgenic Plants

Microprojectile-mediated transformation of rice was carried out based on the procedure described in [Chen, 1998]. Calli were derived from the hypocotyls of germinating mature seeds of the cultivar TP309 (*Oryza sativa*, subspec. *Japonica*). Calli with a diameter of 2-4 mm were selected and placed on a N6 medium (Sigma, Louis, MO, USA) supplemented with 0.3M mannitol and 0.3M sorbitol for about 20 hours before bombardment. Bombardment was carried out with the biolistic PDC-1000/He instrument (Bio-Rad, Hercules, CA, USA). Fifty µl of gold particles (60mg.ml^{-1} at 1:1 ratio of 1.0 and 1.5-3.0 µm diameter gold particle) were coated with effector DNA, target DNA and selection marker DNA in a ratio of 3:3:1 (w/w) and accelerated with a helium pressure of 1100 psi. After two day's incubation, calli were transferred to N6 selection media containing 35mg·l^{-1} hygromycin B and allowed to grow in the dark at 26°C for 45 days. Calli resistant to hygromycin B were transferred to regeneration media and used to generate plantlets as described in [Chen, 1998]. After shoots had reached a height of 1-3 cm, the plantlets were transferred to rooting media (MS plus 0.05 mg·l^{-1} α–Naphthaleneactic acid, Sigma, Louis, MO, USA) and two weeks later the plantlets were transferred to soil and grown in the greenhouse to maturity.

C. PCR Analysis of Transgenic Plants

Genomic DNA was prepared from samples of transgenic plant leaves as described in [Dellaporta, 1983 59] and used as the template for amplification with two pairs of primers for the identification of transgenes. For the rice Reb gene, the forward primer 5'-CCATCCAATCCAATCCACTCCAAC-3' (Fig. 16A) was designed based on a 3' untranslated terminal sequence of the gene, and the reverse primer was designed based on the vector sequence 5'-AGGCGATTAAGTTGGGTAACG-3' (Fig. 16A). For the human lysozyme transgene, the forward primer was designed based on the 5' end of the open reading frame of the gene 5'-CCTAGCCAAAGTCTT CGAGCGGTG-3' (Fig. 16B), and the reverse primer was designed based on the 3' end of the open reading frame of the gene 5'-GCGATGTTGTCTTGCAGC-3' (Fig. 16B). The PCR mixture contained 100 ng genomic DNA, 10 mM Tris (pH 8.3), 1.5 mM $MgCl_2$, 50 mM KCl, and 0.5 μM dNTP. Amplification employed a program of denaturing at 94 °C for 5 min followed by 30 cycles of 94 °C for 30 seconds, 60 °C for 30 seconds and 72 °C for 45 seconds. The PCR products were resolved by electrophoresis in a 1.2 % agarose gel.

D. Lysozyme Activity Assay

A lysozyme assay was carried out using a procedure where twenty individual seeds from each T1 transgenic plant were ground in 1 ml of pre-cold extraction buffer (PBS plus 0.35 M NaCl). After centrifugation at 25,000g for 5 min at 4°C, the supernatant was recovered. A series of dilutions were made and an aliquot was added to a 96-well microtiter plate containing 250 μl of 0.015 % *Micrococcus letus* cells in each well (Sigma, Louis, MO, USA; procedure developed at Applied Phytologics Inc.). Human lysozyme (*EC 3.2.1.17*, Sigma, Louis, MO, USA) was used as the standard and lysozyme activity was measured based on the decrease in turbidity, evaluated using a Microplate Reader 3550 (Bio-Rad, Hercules, CA, USA). The lysozyme concentration in the samples was determined based on absorbance values of samples relative to a standard curve prepared using different concentrations of human lysozyme. The lysozyme expression level in a given transgenic plant was calculated as the average lysozyme content of the twenty seeds taken from that plant. Total soluble protein in seed extracts was estimated using the Bradford protein assay (Bio-Rad, Hercules, CA, USA).

E. Enhanced Human Lysozyme Expression in Transgenic Rice Seed Co-Transformed with Reb

In order to evaluate the ability of Reb to increase the expression of a transgene in transgenic plants, plants cells were co-transformed with heterologous nucleic acid constructs comprising the human lysozyme gene driven by the Glb promoter (Glb-lys) and the Reb gene driven by the actin promoter (Act-Reb) and with the Act-Reb construct alone. Normal plant phenotypes were obtained among transformants containing both Glb-lys and native Reb and PCR analysis confirmed that both genes had integrated into the genome in 10 out of 11 plants (Fig. 17). Transgenic seeds were harvested at maturity and lysozyme activity was analyzed.

Lysozyme expression in seeds of the 10 transgenic plants containing native-Reb and Glb-lys ranged from 31-133 $\mu g \cdot mg^{-1}$ soluble protein with an average of 69.8±11.6 $\mu g \cdot mg^{-1}$ soluble protein (Fig. 17). Seeds taken from seventeen transgenic plants containing Glb-Lys alone expressed lysozyme in amounts ranging from 7 to 76 $\mu g \cdot mg^{-1}$ soluble protein with an average of 33.95±4.96 $\mu g \cdot mg^{-1}$ soluble protein (Fig. 17). No lysozyme activity was detected in untransformed rice seeds (Fig. 17). The results showed that the expression level of the lysozyme increased on an average 2-fold, when the seeds were transgenic for both the Reb gene and Glb-lys. The statistical analysis (t-test) showed that the amount of lysozyme in seeds from the plants transgenic for the Reb gene and Glb-lys is significantly higher than in the plants with Glb-lys alone ($P<0.001$).

Brief Description of the Sequences

Description	SEQ ID NO
Codon optimized lysozyme coding sequence: AAAGTCTTCGAGCGGTGCGAGCTGGCCCGCACGCTCAAGCGGCTCGGCAT GGACGGCTACCGGGGCATCAGCCTCGCCAACTGGATGTGCCTCGCCAAGT GGGAGTCGGGCTACAACACCCGCGCAACCAACTACAACGCCGGCGACCGC TCCACCGACTACGGCATCTTCCAGATCAACTCCCGCTACTGGTGCAACGAC GGCAAGACGCCCGGGGCCGTCAACGCCTGCCACCTCTCCTGCTCGGCCCT GCTGCAAGACAACATCGCCGACGCCGTCGCGTGCGCGAAGCGCGTCGTCC GCGACCCGCAGGGCATCCGGGCCTGGGTGGCCTGGCGCAACCGCTGCCA GAACCGGGACGTGCGCCAGTACGTCCAGGGCTGCGGCGTCTGA	1
Amino acid sequence based on codon optimized lysozyme coding sequence: KVFERCELARTLKRLGMDGYRGISLANWMCLAKWESGYNTRATNYNAGDRST DYGIFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVRDPQGI RAWVAWRNRCQNRDVRQYVQGCGV	2
Codon optimized lactoferrin coding sequence:GGGCGGCGGCGGCGCTCGGTGCAGTGGTGCGCCGTGTCCCAGC CCGAGGCGACCAAGTGCTTCCAGTGGCAGCGCAACATGCGGAAGGTGCGC GGCCCGCCGGTCAGCTGCATCAAGCGGGACTCCCCCATCCAATGCATCCAG GCCATCGCGGAGAACCGCGCCGACGCGGTCACCCTGGACGGCGGGTTCAT CTACGAGGCGGGGCTCGCCCCGTACAAGCTCCGCCCGGTGGCGGCGGAG GTGTACGGCACCGAGCGCCAGCCGCGCACGCACTACTACGCGGTGGCCGT CGTCAAGAAGGGCGGGTCCTTCCAGCTCAACGAGCTGCAGGGCCTGAAGT CGTGCCACACGGGCCTCCGGCGGACGGCGGGCTGGAACGTGCCCATCGG CACCCTGCGCCCCTTCCTGAACTGGACCGGCCCGCCGGAGCCGATCGAGG CCGCCGTGGCCCGCTTCTTCAGCGCCTCCTGCGTCCCCGGCGCCGACAAG GGCCAGTTCCCGAACCTCTGCCGGCTCTGCGCCGGGACGGGCGAGAACAA GTGCGCCTTCTCCTCGCAGGAGCCGTACTTCTCCTACTCGGGCGCGTTCAA GTGCCTCCGCGACGGGGCCGGCGACGTGGCGTTCATCCGCGAGTCCACCG TGTTCGAGGACCTCTCCGACGAGGCGGAGCGGGACGAGTACGAGCTGCTG TGCCCCGACAACACCCGCAAGCCGGTGGACAAGTTCAAGGACTGCCACCTG GCGCGGGTGCCCTCGCACGCGGTCGTCGCCCGCAGCGTCAACGGCAAGGA GGACGCGATCTGGAACCTCCTCCGCCAGGCCCAGGAGAAGTTCGGCAAGG ACAAGTCCCCCAAGTTCCAGCTCTTCGGGAGCCCCAGCGGCCAGAAGGACC TCCTCTTCAAGGACTCCGCGATCGGCTTCTCCCGCGTCCCCCCGCGCATCG ACTCCGGCCTGTACCTCGGCTCCGGGTACTTCACCGCGATCCAGAACCTCC GGAAGAGCGAGGAGGAGGTGGCGGCGCGGCGGGCCCGCGTCGTGTGGTG CGCCGTGGGCGAGCAGGAGCTGCGGAAGTGCAACCAGTGGAGCGGCCTGA GCGAGGGGTCGGTGACCTGCTCGTCCGCCAGCACCACCGAGGACTGCATC GCGCTCGTCCTCAAGGGGGAGGCCGACGCGATGAGCCTCGACGGGGGGTA CGTCTACACCGCCGGCAAGTGCGGCCTGGTCCCGGTCCTGGCGGAGAACT ACAAGTCGCAGCAGTCCAGCGACCCCGACCCGAACTGCGTGGACCGCCCC GTCGAGGGCTACCTCGCCGTGGCCGTCGTGCGCCGGTCCGACACCTCCCT GACGTGGAACAGCGTCAAGGGCAAGAAGAGCTGCCACACCGCCGTGGACC GCACCGCCGGCTGGAACATCCCGATGGGCCTCCTCTTCAACCAGACCGGCT CCTGCAAGTTCGACGAGTACTTCTCCCAGTCCTGCGCCCCCGGCTCGGACC CCCGCTCCAACCTGTGCGCCCTCTGCATCGGGGACGAGCAGGGCGAGAAC AAGTGCGTGCCCAACAGCAACGAGCGGTACTACGGCTACACGGGGGCCTT CCGCTGCCTGGCGGAGAACGCCGGGGACGTCGCGTTCGTGAAGGACGTGA CCGTGCTGCAAAACACGGACGGGAACAACAACGAGGCGTGGGCGAAGGAC CTCAAGCTCGCCGACTTCGCCCTGCTGTGCCTCGACGGCAAGCGCAAGCCC GTCACCGAGGCGCGGTCCTGCCACCTGGCGATGGCCCCCAACCACGCCGT CGTCTCCCGCATGGACAAGGTCGAGCGCCTCAAGCAGGTGCTCCTGCACCA GCAGGCCAAGTTCGGCCGGAACGGCAGCGACTGCCCGGACAAGTTCTGCC	3

TGTTCCAGTCGGAGACCAAGAACCTCCTCTTCAACGACAACACCGAGTGCCT GGCGCGCCTCCACGGCAAGACCACCTACGAGAAGTACCTCGGCCCGCAGT ACGTCGCCGGCATCACCAACCTCAAGAAGTGCTCCACCTCCCCCCTCCTGG AGGCGTGCGAGTTCCTCCGCAAGTGA	
Amino acid sequence based on codon optimized lactoferrin coding sequence: GRRRRSVQWCAVSQPEATKCFQWQRNMRKVRGPPVSCIKRDSPIQCIQAIAEN RADAVTLDGGFIYEAGLAPYKLRPVAAEVYGTERQPRTHYYAVAVVKKGGSFQL NELQGLKSCHTGLRRTAGWNVPIGTLRPFLNWTGPPEPIEAAVARFFSASCVPG ADKGQFPNLCRLCAGTGENKCAFSSQEPYFSYSGAFKCLRDGAGDVAFIREST VFEDLSDEAERDEYELLCPDNTRKPVDKFKDCHLARVPSHAVVARSVNGKEDAI WNLLRQAQEKFGKDKSPKFQLFGSPSGQKDLLFKDSAIGFSRVPPRIDSGLYLG SGYFTAIQNLRKSEEEVAARRARVVWCAVGEQELRKCNQWSGLSEGSVTCSS ASTTEDCIALVLKGEADAMSLDGGYVYTAGKCGLVPVLAENYKSQQSSDPDPN CVDRPVEGYLAVAVVRRSDTSLTWNSVKGKKSCHTAVDRTAGWNIPMGLLFNQ TGSCKFDEYFSQSCAPGSDPRSNLCALCIGDEQGENKCVPNSNERYYGYTGAF RCLAENAGDVAFVKDVTVLQNTDGNNNEAWAKDLKLADFALLCLDGKRKPVTE ARSCHLAMAPNHAVVSRMDKVERLKQVLLHQQAKFGRNGSDCPDKFCLFQSE TKNLLFNDNTECLARLHGKTTYEKYLGPQYVAGITNLKKCSTSPLLEACEFLRK	4
MV-Gt1-F1 primer: 5' ATC GAA GCT TCA TGA GTA ATG TGT GAG CAT TAT GGG ACC ACG 3'	5
Xba-Gt1-R1 primer: 5' CTA GTC TAG ACT CGA GCC ATG GGG CCG GCT AGG GAG CCA TCG CAC AAG AGG AA 3'	6
Codon optimized lactoferricin coding sequence ACCAAGTGCTTCCAGTGGCAGCGCAACATGCGGAAGGTGCGCGGCCCGCC GGTCAGCTGCATCAAGCGGGAC	7
Codon optimized EGF coding sequence AACTCCGACTCGGAGTGCCCCCTCTCCCACGACGGTTACTGCCTCCACGAC GGGGTCTGCATGTACATCGAGGCCCTCGACAAGTACGCCTGCAACTGCGTC GTGGGCTACATCGGCGAGCGGTGCCAGTACCGCGACCTCAAGTGGTGGGA GCTGCGCTGA	8
Codon optimized IGF-1 coding sequence GGCCCGGAGACCCTCTGCGGCGCCGAGCTCGTGGACGCCCTCCAGTTCGT GTGCGGCGACCGCGGCTTCTACTTCAACAAGCCGACCGGCTACGGCAGCA GCAGCCGCCGCGCCCCGCAGACCGGCATCGTGGACGAGTGCTGCTTCCGC AGCTGCGACCTCCGCCGCCTGGAGATGTACTGCGCCCCGCTCAAGCCCGC CAAGAGCGCCTGA	9
Codon optimized lactohedrin coding sequence CTGGACATCTGCTCGAAGAACCCGTGCCACAACGGCGGGCTCTGCGAGGA GATCAGCCAGGAGGTGCGGGGCGACGTGTTCCCCTCGTACACCTGCACCT GCCTGAAGGGCTACGCCGGGAACCACTGCGAGACGAAGTGCGTGGAGCCC CTGGGGATGGAGAACGGCAACATCGCCAACTCCCAGATCGCCGCCTCCTCC GTGCGGGTGACCTTCCTCGGCCTCCAGCACTGGGTCCCGGAGCTGGCCCG GCTCAACCGGGCGGGCATGGTGAACGCGTGGACCCCCTCGTCCAACGACG ACAACCCGTGGATCCAAGTGAACCTGCTCCGCCGCATGTGGGTCACCGGCG TGGTCACCCAAGGCGCCAGCCGCCTGGCCAGCCACGAGTACCTCAAGGCC TTCAAGGTCGCCTACAGCCTCAACGGCCACGAGTTCGACTTCATCCACGAC GTCAACAAGAAGCACAAGGAGTTCGTGGGCAACTGGAACAAGAACGCGGTC CACGTGAACCTCTTCGAGACCCCCGTCGAGGCCCAGTACGTCCGCCTCTAC CCCACGAGCTGCCACACCGCCTGCACGCTCCGCTTCGAGCTGCTGGGGTG CGAGCTGAACGGGTGCGCGAACCCGCTGGGGCTCAAGAACAACAGCATCC CCGACAAGCAGATCACGGCCTCGTCGTCGTACAAGACCTGGGGCCTGCACC TCTTCTCGTGGAACCCGAGCTACGCCCGGCTGGACAAGCAGGGCAACTTCA ACGCCTGGGTCGCCGGGAGCTACGGGAACGACCAGTGGCTCCAGGTGGAC CTCGGCAGCTCCAAGGAGGTCACCGGCATCATCACGCAGGGGGCCCGCAA CTTCGGCTCCGTGCAGTTCGTGGCCTCCTACAAGGTGGCCTACTCGAACGA	10

CAGCGCCAACTGGACCGAGTACCAGGACCCGCGCACCGGGTCCAGCAAGA TCTTCCCCGGCAACTGGGACAACCACAGCCACAAGAAGAACCTGTTCGAGA CCCCCATCCTCGCCCGGTACGTCCGCATCCTCCCCGTCGCTTGGCACAACC GGATCGCGCTCCGGCTGGAGCTCCTCGGCTGCTGA	
Codon optimized kappa-casein coding sequence GAGGTCCAAAACCAGAAGCAGCCCGCCTGCCACGAGAACGACGAGCGCCC CTTCTACCAGAAGACCGCACCCTACGTCCCGATGTACTACGTCCCGAACAG CTACCCCTACTACGGTACGAACCTGTACCAGCGCCGCCCGGCCATCGCTAT CAACAACCCCTACGTCCCCCGGACCTACTACGCGAACCCGGCCGTGGTGC GGCCCCACGCGCAGATCCCGCAGCGGCAGTACCTGCCAAACAGCCACCCC CCCACCGTGGTGCGGCGGCCCAACCTCCACCCGAGCTTCATCGCTATCCCC CCCAAGAAGATCCAGGACAAGATCATCATCCCGACCATCAACACCATCGCCA CCGTGGAGCCGACGCCAGCCCCCGCGACCGAGCCCACGGTGGACAGCGT CGTGACCCCAGAGGCGTTCTCCGAATCGATCATCACCTCCACCCCCGAGAC CACCACGGTGGCCGTCACGCCGCCGACGGCATGA	11
Codon optimized haptocorrin coding sequence GAGATCTGCGAGGTCTCCGAGGAGAACTACATCCGCCTCAAGCCCCTCCTG AACACCATGATCCAGAGCAACTACAACCGGGGCACGTCGGCCGTGAACGTC GTGCTCTCCCTGAAGCTCGTGGGCATCCAGATCCAGACCCTCATGCAGAAG ATGATCCAGCAGATCAAGTACAACGTGAAGAGCCGCCTCTCGGACGTGTCC AGCGGCGAGCTGGCGCTCATCATCCTCGCGCTCGGCGTGTGCCGGAACGC GGAGGAGAACCTCATCTACGACTACCACCTCACGGACAAGCTGGAGAACAA GTTCCAGGCCGAGATCGAGAACATGGAGGCCCACAACGGCACCCCGCTGA CCAACTACTACCAGCTCAGCCTGGACGTCCTCGCGCTCTGCCTGTTCAACG GGAACTACTCCACCGCCGAGGTGGTCAACCACTTCACCCCCGAGAACAAGA ACTACTACTTCGGCTCGCAGTTCTCCGTGGACACCGGGGCCATGGCCGTCC TGGCCCTCACCTGCGTGAAGAAGTCCCTCATCAACGGCCAGATCAAGGCCG ACGAGGGCTCCCTGAAGAACATCTCGATCTACACCAAGAGCCTCGTGGAGA AGATCCTCAGCGAGAAGAAGGAGAACGGGCTGATCGGCAACACCTTCTCGA CCGGCGAGGCGATGCAGGCCCTGTTCGTGAGCAGCGACTACTACAACGAG AACGACTGGAACTGCCAGCAGACCCTCAACACGGTCCTGACCGAGATCAGC CAGGGCGCGTTCAGCAACCCCAACGCCGCCGCCCAGGTCCTGCCGGCCCT GATGGGCAAGACCTTCCTCGACATCAACAAGGACAGCTCCTGCGTGTCCGC GAGCGGCAACTTCAACATCTCCGCCGACGAGCCGATCACGGTGACGCCGC CCGACAGCCAGTCGTACATCTCCGTGAACTACAGCGTGCGGATCAACGAGA CCTACTTCACGAACGTGACGGTCCTCAACGGCTCGGTCTTCCTGAGCGTGA TGGAGAAGGCGCAGAAGATGAACGACACGATCTTCGGCTTCACGATGGAGG AGCGCAGCTGGGGCCCCTACATCACCTGCATCCAGGGCCTCTGCGCCAACA ACAACGACCGCACCTACTGGGAGCTGCTGAGCGGCGGCGAGCCGCTGAGC CAGGGGGCCGGCAGCTACGTGGTCCGCAACGGCGAGAACCTGGAGGTCCG GTGGAGCAAGTACTGA	12
Codon optimized lactoperoxidase coding sequence CAAACGACCCGGACGTCGGCGATCTCCGACACGGTCTCGCAGGCCAAGGT GCAAGTCAACAAGGCATTCCTGGATTCGCGCACGCGGCTGAAGACCGCGAT GTCGTCCGAGACCCCGACGAGCCGGCAGCTGAGCGAGTACCTCAAGCACG CGAAGGGGCGGACGCGCACCGCCATCCGCAATGGCCAAGTGTGGGAGGAA TCCCTGAAGCGGCTGCGGCAGAAGGCGTCGCTCACCAACGTGACCGACCC GTCCCTCGACCTGACCAGCCTCTCCCTGGAGGTCGGCTGCGGCGCCCCGG CGCCCGTCGTGCGCTGCGACCCCTGCTCGCCATACCGCACGATCACGGGC GACTGCAACAACCGGCGGAAGCCGGCACTGGGGGCTGCGAACCGCGCCCT CGCGCGCTGGCTCCCCGCCGAGTACGAGGACGGCCTCAGCCTCCCCTTCG GTTGGACCCCCGGCAAGACGCGCAACGGCTTCCCGCTCCCGCTCGCTCGC GAGGTCAGCAACAAGATCGTCGGTTACCTGAACGAGGAGGGGGTCCTCGAC CAAAACCGCTCCCTCCTCTTCATGCAGTGGGGGCAGATCGTGGACCACGAC CTGGACTTCGCCCCGGACACGGAGCTGGGCTCCAGCGAGTACAGCAAGAC CCAGTGCGACGAATACTGCATCCAGGGCGACAACTGCTTCCCGATCATGTT CCCCCCGAACGACCCGAAGGCGGGCACCCAGGGCAAGTGCATGCCGTTCT TCCGGGCAGGCTTCGTCTGCCCGACCCCCCCGTACAAGTCCCTCGCGCGC	13

GAGCAGATCAACGCGCTCACGTCCTTCCTCGACGCCAGCTTCGTCTACAGC AGCGAGCCGTCCCTCGCCAGCCGCCTCCGCAACCTCAGCAGCCCCCTCGG CCTCATGGCGGTCAACCAGGAGGTGTCGGACCACGGCCTCCCATACCTGCC GTACGACAGCAAGAAGCCGTCCCCCTGCGAGTTCATCAACACCACCGCGCG CGTCCCGTGCTTCCTCGCCGGCGATTCGCGGGCGAGCGAGCACATCCTCC TCGCCACGAGCCACACCCTGTTCCTCCGCGAGCACAACCGCCTCGCCCGG GAGCTGAAGCGCCTCAACCCGCAGTGGGACGGCGAGAAGCTCTACCAGGA GGCCCGGAAGATCCTCGGCGCTTTCGTCCAGATCATCACCTTCCGGGACTA CCTCCCCATCCTGCTCGGTGACCACATGCAGAAGTGGATCCCCCCCTACCA AGGCTACTCCGAGAGCGTGGACCCGCGCATCTCCAACGTCTTCACGTTCGC GTTCCGCTTCGGGCACCTGGAGGTGCCGTCGTCGATGTTCCGCCTCGACGA GAACTACCAGCCCTGGGGCCCAGAGCCGGAGCTGCCGCTCCACACCCTGT TCTTCAACACCTGGCGGATGGTCAAGGACGGCGGCATCGACCCGCTCGTGC GCGGGCTCCTGGCTAAGAAGTCGAAGCTCATGAAGCAGAACAAGATGATGA CCGGCGAGCTGCGCAACAAGCTGTTCCAGCCCACCCACCGCATCCACGGG TTCGACCTGGCTGCAATCAACACCCAGCGGTGCCGCGACCACGGCCAGCC CGGCTACAACTCGTGGCGCGCGTTCTGCGACCTCTCCCAGCCACAGACGCT GGAGGAGCTCAACACCGTGCTCAAGAGCAAGATGCTCGCCAAGAAGCTGCT CGGGCTCTACGGCACGCCCGACAACATCGACATCTGGATCGGGGCCATCG CGGAGCCGCTCGTGGAGCGCGGGCGCGTCGGCCCGCTGCTCGCGTGCCT CCTGGGCAAGCAATTCCAACAGATCCGCGACGGGGACCGGTTCTGGTGGG AGAACCCCGGCGTGTTCACCAACGAGCAGAAGGATTCGCTCCAAAAGATGA GCTTCTCCCGCCTGGTGTGCGACAACACCCGCATCACCAAGGTCCCGCGCG ACCCATTCTGGGCCAACTCCTACCCGTACGACTTCGTGGACTGCTCCGCCA TCGACAAGCTCGACCTGTCCCCCTGGGCATCGGTGAAGAACTGA	
Codon optimized alpha-1-antitrypsin coding sequence GAGGACCCGCAGGGCGACGCCGCCCAGAAGACCGACACCAGCCACCACGA CCAGGACCACCCGACGTTCAACAAGATCACCCCGAATTTGGCCGAATTCGC CTTCAGCCTGTACCGCCAGCTCGCGCACCAGTCCAACTCCACCAACATCTTC TTCAGCCCGGTGAGCATCGCCACCGCCTTCGCCATGCTGTCCCTGGGTACC AAGGCGGACACCCACGACGAGATCCTCGAAGGGCTGAACTTCAACCTGACG GAGATCCCGGAGGCGCAGATCCACGAGGGCTTCCAGGAGCTGCTCAGGAC GCTCAACCAGCCGGACTCCCAGCTCCAGCTCACCACCGGCAACGGGCTCTT CCTGTCCGAGGGCCTCAAGCTCGTCGATAAGTTCCTGGAGGACGTGAAGAA GCTCTACCACTCCGAGGCGTTCACCGTCAACTTCGGGGACACCGAGGAGGC CAAGAAGCAGATCAACGACTACGTCGAGAAGGGGACCCAGGGCAAGATCGT GGACCTGGTCAAGGAATTGGACAGGGACACCGTCTTCGCGCTCGTCAACTA CATCTTCTTCAAGGGCAAGTGGGAGCGCCCGTTCGAGGTGAAGGACACCGA GGAGGAGGACTTCCACGTCGACCAGGTCACCACCGTCAAGGTCCCGATGAT GAAGAGGCTCGGCATGTTCAACATCCAGCACTGCAAGAAGCTCTCCAGCTG GGTGCTCCTCATGAAGTACCTGGGGAACGCCACCGCCATCTTCTTCCTGCC GGACGAGGGCAAGCTCCAGCACCTGGAGAACGAGCTGACGCACGACATCA TCACGAAGTTCCTGGAGAACGAGGACAGGCGCTCCGCTAGCCTCCACCTCC CGAAGCTGAGCATCACCGGCACGTACGACCTGAAGAGCGTGCTGGGCCAG CTGGGCATCACGAAGGTCTTCAGCAACGGCGCGGACCTCTCCGGCGTGAC GGAGGAGGCCCCCCTGAAGCTCTCCAAGGCCGTGCACAAGGCGGTGCTCA CGATCGACGAGAAGGGGACGGAAGCTGCCGGGGCCATGTTCCTGGAGGCC ATCCCCGTGTCCATCCCGCCCGAGGTCAAGTTCAACAAGCCCTTCGTCTTCC TGATGATCGAGCAGAACACGAAGAGCCCCCTCTTCATGGGGAAGGTCGTCA ACCCCACGCAGAAGTGA Codon optimized immunoglobulin-A coding sequence	14
Rice Gt1 promoter and Gt1 leader coding sequence CATGAGTAATGTGTGAGCATTATGGGACCACGAAATAAAAAGAACATTTTGAT GAGTCGTGTATCCTCGATGAGCCTCAAAAGTTCTCTCACCCCGGATAAGAAA CCCTTAAGCAATGTGCAAAGTTTGCATTCTCCACTGACATAATGCAAAATAAG ATATCATCGATGACATAGCAACTCATGCATCATATCATGCCTCTCTCAACCTA TTCATTCCTACTCATCTACATAAGTATCTTCAGCTAAATGTTAGAACATAAACC	15

CATAAGTCACGTTTGATGAGTATTAGGCGTGACACATGACAAATCACAGACT CAAGCAAGATAAAGCAAAATGATGTGTACATAAAACTCCAGAGCTATATGTCA TATTGCAAAAAGAGGAGAGCTTATAAGACAAGGCATGACTCACAAAAATTCA CTTGCCTTTCGTGTCAAAAAGAGGAGGGCTTTACATTATCCATGTCATATTGC AAAAGAAAGAGAGAAAGAACAACACAATGCTGCGTCAATTATACATATCTGTA TGTCCATCATTATTCATCCACCTTTCGTGTACCACACTTCATATATCATAAGA GTCACTTCACGTCTGGACATTAACAAACTCTATCTTAACATTTAGATGCAAGA GCCTTTATCTCACTATAAATGCACGATGATTTCTCATTGTTTCTCACAAAAAG CGGCCGCTTCATTAGTCCTACAACAACATGGCATCCATAAATCGCCCCATAG TTTTCTTCACAGTTTGCTTGTTCCTCTTGTGCGATGGCTCCCTAGCC Codon optimized alpha-1-antitrypsin coding sequence	
Rice Glb promoter and Gt1 leader coding sequence CTGCAGGGAGGAGAGGGGAGAGATGGTGAGAGAGGAGGAAGAAGAGGAG GGGTGACAATGATATGTGGGGCATGTGGGCACCCAATTTTTTAATTCATTCT TTTGTTGAAACTGACATGTGGGTCCCATGAGATTTATTATTTTTCGGATCGAA TCGCCACGTAAGCGCTACGTCAATGCTACGTCAGATGAAGACCGAGTCAAAT TAGCCACGTAAGCGCCACGTCAGCCAAAACCACCATCCAAACCGCCGAGGG ACCTCATCTGCACTGGTTTTGATAGTTGAGGGACCCGTTGTATCTGGTTTTTC GATTGAAGGACGAAAATCAAATTTGTTGACAAGTTAAGGGACCTTAAATGAA CTTATTCCATTTCAAAATATTCTGTGAGCCATATATACCGTGGGCTTCCAATC CTCCTCAAATTAAAGGGCCTTTTTAAAATAGATAATTGCCTTCTTTCAGTCAC CCATAAAAGTACAAAACTACTACCAACAAGCAACATGCGCAGTTACACACATT TTCTGCACATTTCCGCCACGTCACAAAGAGCTAAGAGTTATCCCTAGGACAA TCTCATTAGTGTAGATACATCCATTAATCTTTTATCAGAGGCAAACGTAAAGC CGCTCTTTATGACAAAAATAGGTGACACAAAAGTGTTATCTGCCACATACATA ACTTCAGAAATTACCCAACACCAAGAGAAAAATAAAAAAAAAATCTTTTTGCAA GCTCCAAATCTTGGAAACCTTTTTCACTCTTTGCAGCATTGTACTCTTGCTCT TTTTCCAACCGATCCATGTCACCCTCAAGCTTCTACTTGATCTACACGAAGCT CACCGTGCACACAACCATGGCCACAAAAACCCTATAAAACCCCATCCGATCG CCATCATCTCATCATCAGTTCATTACCAACAAACAAAAGAGGAAAAAAAACAT ATACACTTCTAGTGATTGTCTGATTGATCATCAATCTAGAGGCGGCCGCATG GCTAGCAAGGTCGTCTTCTTCGCGGCGGCGCTCATGGCGGCCATGGTGGC CATCTCCGGC Rice Gt1 promoter and Gt1 leader coding sequence	16
Other monocot maturation specific promoter sequences Seq #17 Bx7 promoter seq CTGCAGGCCAGGGAAAGACAATGGACATGCAAAGAGGTAGGGGCAGGGAA GAAACACTTGGAGATCATAGAAGAACATAAGAGGTTAAACATAGGAGGGCAT AATGGACAATTAAATCTACATTAATTGAACTCATTTGGGAAGTAAACAAAATC CATATTCTGGTGTAAATCAAACTATTTGACGCGGATTTACTAAGATCCTATGT TAATTTTAGACATGACTGGCCAAAGGTTTCAGTTAGTTCATTTGTCACGGAAA GGTGTTTTCATAAGTCCAAAACTCTACCAACTTTTTTGCACGTCATAGCATAG ATAGATGTTGTGAGTCATTGGATAGATATTGTGAGTCAGCATGGATTTGTGTT GCCTGGAAATCCAACTAAATGACAAGCAACAAAACCTGAAATGGGCTTTAGG AGAGATGGTTTATCAATTTACATGTTCCATGCAGGCTACCTTCCACTACTCGA CATGGTTAGAAGTTTTGAGTGCCGCATATTTGCGGAAGCAATGGCACTACTC GACATGGTTAGAAGTTTTGAGTGCCGCATATTTGCGGAAGCAATGGCTAACA GATACATATTCTGCCAAACCCCAAGAAGGATAATCACTCCTCTTAGATAAAAA GAACAGACCAATGTACAAACATCCACACTTCTGCAAACAATACACCAGAACT AGGATTAAGCCCATTACGTGGCTTTAGCAGACCGTCCAAAAATCTGTTTTGC AAGCACCAATTGCTCCTTACTTATCCAGCTTCTTTTGTGTTGGCAAACTGCCC TTTTCCAACCGATTTTGTTTCTTCTCACGCTTTCTTCATAGGCTAAACTAACCT CGGCGTGCACACAACCATGTCCTGAACCTTCACCTCGTCCCTATAAAAGCCC ATCCAACCTTACAATCTCATCATCACCCACAACACCGAGCACCCCAATCTAC AGATCAATTCACTGACAGTTCACTGATCTAGA	17-23

Seq #18 Glub-2 promoter seq

CTGCAGTAATGGATACCTAGTAGCAAGCTAGCTTAAACAAATCTAAATTCCAA
TCTGTTCGTAAACGTTTTCTCGATCGCAATTTTGATCAAAACTATTGAAAACC
TCAATTAAACCATTCAAAATTTTTAATATACCCAACAAGAGCGTCCAAACCAA
ATATGTAAATATGGATGTCATGATAATTGACTTATGACAATGTGATTATTTCAT
CAAGTCTTTAAATCATTAATTCTAGTTGAAGGTTTATGTTTTCTTATGCTAAAG
GGTTATGTTTATATAAGAATATTAAAGAGCAAATTGCAATAGATCAACACAAC
AAATTTGAATGTTTCCAGATGTGTAAAAATATCCAAATTAATTGTTTTAAAATA
GTTTTAAGAAGGATCTGATATGCAAGTTTGATAGTTAGTAAACTGCAAAAGGG
CTTATTACATGGAAAATTCCTTATTGAATATGTTTCATTGACTGGTTTATTTTA
CATGACAACAAAGTTACTAGTATGTCAATAAAAAAATACAAGGTTACTTGTCA
ATTGTATTGTGCCAAGTAAAGATGACAACAAACATACAAATTTATTTGTTCTTT
TATAGAAACACCTAACTTATCAAGGATAGTTGGCCACGCAAAAATGACAACAT
ACTTTACAATTGTATCATCATAAAGATCTTATCAAGTATAAGAACTTTATGGTG
ACATAAAAAAATAATCACAAGGGCAAGACACATACTAAAAGTATGGACAGAAAT
TTCTTAACAAACTCCATTTGTTTTGTATCCAAAAGCATAAGAAATGAGTCATG
GCTGAGTCATGATATGTAGTTCAATCTTGCAAAATTGCCTTTTTGTTAAGTATT
GTTTTAACACTACAAGTCACATATTGTCTATACTTGCAACAAACACTATTACC
GTGTATCCCAAGTGGCCTTTTCATTGCTATATAAACTAGCTTGATCGGTCTTT
CAACTCACATCAATTAGCTTAAGTTTCCATTAGCAACTGCTAATAGCT

Seq#19 Gt3 promoter seq

CTGCAGTGTAAGTGTAGCTTCTTATAGCTTAGTGCTTTACTATCTTCACAAGC
ACATGCTATAGTATTGTTCCAAGATGAAAGAATAATTCATCCTTGCTACCAAC
TTGCATGATATTATATTTGTGAATATCCTATCTCTTGGCTTATAATGAAATGTG
CTGCTGGGTTATTCTGACCATGGTATTTGAGAGCCTTTGTATAGCTGAAACC
AACGTATATCGAGCATGGAACAGAGAACAAAATGCAAGGATTTTTTTATTCTG
GTTCATGCCCTGGATGGGTTAATATCGTGATCATCAAAAAAGATATGCATAAA
ATTAAAGTAATAAATTTGCTCATAAGAAACCAAAACCAAAAGCACATATGTCC
TAAACAAACTGCATTTTGTTTGTCATGTAGCAATACAAGAGATAATATATGAC
GTGGTTATGACTTATTCACTTTTTGTGACTCCAAAATGTAGTAGGTCTAACTG
ATTGTTTAAAGTGATGTCTTACTGTAGAAGTTTCATCCCAAAAGCAATCACTA
AAGCAACACACACGTATAGTCCACCTTCACGTAATTCTTTGTGGAAGATAACA
AGAAGGCTCACTGAAAAATAAAAGCAAAGAAAAGGATATCAAACAGACCATT
GTGCATCCCATTGATCCTTGTATGTCTATTTATCTATCCTCCTTTTGTGTACCT
TACTTCTATCTAGTGAGTCACTTCATATGTGGACATTAACAAACTCTATCTTAA
CATCTAGTCGATCACTACTTTACTTCACTATAAAAGGACCAACATATATCATC
CATTTCTCACAAAAGCATTGAGTTCAGTCCCACAAAATCTAGA

Seq #20 Glub-1 promoter seq

CTGCAGAGATATGGATTTTCTAAGATTAATTGATTCTCTGTCTAAAGAAAAAA
AGTATTATTGAATTAAATGGAAAAAGAAAAAGGAAAAAGGGGATGGCTTCTG
CTTTTTGGGCTGAAGGCGGCGTGTGGCCAGCGTGCTGCGTGCGGACAGCG
AGCGAACACACGACGGAGCAGCTACGACGAACGGGGGACCGAGTGGACCG
GACGAGGATGTGGCCTAGGACGAGTGCACAAGGCTAGTGGACTCGGTCCC
CGCGCGGTATCCCGAGTGGTCCACTGTCTGCAAACACGATTCACATAGAGC
GGGCAGACGCGGGAGCCGTCCTAGGTGCACCGGAAGCAAATCCGTCGCCT
GGGTGGATTTGAGTGACACGGCCCACGTGTAGCCTCACAGCTCTCCGTGGT
CAGATGTGTAAAATTATCATAATATGTGTTTTTCAAATAGTTAAATAATATATAT
AGGCAAGTTATATGGGTCAATAAGCAGTAAAAAGGCTTATGACATGGTAAAA
TTACTTACACCAATATGCCTTACTGTCTGATATATTTTACATGACAACAAAGTT
ACAAGTACGTCATTTAAAAATACAAGTTACTTATCAATTGTAGTGTATCAAGTA
AATGACAACAAACCTACAAATTTGCTATTTTGAAGGAACACTTAAAAAAATCA
ATAGGCAAGTTATATAGTCAATAAACTGCAAGAAGGCTTATGACATGGAAAAA
TTACATACACCAATATGCTTTATTGTCCGGTATATTTTACAAGACAACAAAGTT

ATAAGTATGTCATTTAAAAATACAAGTTACTTATCAATTGTCAAGTAAATGAAA
ACAAACCTACAAATTTGTTATTTTGAAGGAACACCTAAATTATCAAATATAGCT
TGCTACGCAAAATGACAACATGCTTACAAGTTATTATCATCTTAAAGTTAGAC
TCATCTTCTCAAGCATAAGAGCTTTATGGTGCAAAAACAAATATAATGACAAG
GCAAAGATACATACATATTAAGAGTATGGACAGACATTTCTTTAACAAACTCC
ATTTGTATTACTCCAAAAGCACCAGAAGTTTGTCATGGCTGAGTCATGAAATG
TATAGTTCAATCTTGCAAAGTTGCCTTTCCTTTTGTACTGTGTTTTAACACTAC
AAGCCATATATTGTCTGTACGTGCAACAAACTATATCACCATGTATCCCAAGA
TGCTTTTTTATTGCTATATAAACTAGCTTGGTCTGTCTTTGAACTCACATCAAT
TAGCTTAAGTTTCCATAAGCAAGTACAAATAGCTCTAGA

Seq #21 Rice proalmin promoter seq

CTGCAGCATCGGCTTAGGTGTAGCAACACGACTTTATTATTATTATTATTATT
ATTATTATTATTTTACAAAAATATAAAATAGATCAGTCCCTCACCACAAGTAGA
GCAAGTTGGTGAGTTATTGTAAAGTTCTACAAAGCTAATTTAAAAGTTATTGC
ATTAACTTATTTCATATTACAAACAAGAGTGTCAATGGAACAATGAAAACCAT
ATGACATACTATAATTTTGTTTTTATTATTGAAATTATATAATTCAAAGAGAATA
AATCCACATAGCCGTAAAGTTCTACATGTGGTGCATTACCAAAATATATATAG
CTTACAAAACATGACAAGCTTAGTTTGAAAAATTGCAATCCTTATCACATTGA
CACATAAAGTGAGTGATGAGTCATAATATTATTTTTCTTGCTACCCATCATGT
ATATATGATAGCCACAAAGTTACTTTGATGATGATATCAAAGAACATTTTTAG
GTGCACCTAACAGAATATCCAAATAATATGACTCACTTAGATCATAATAGAGC
ATCAAGTAAAACTAACACTCTAAAGCAACCGATGGGAAAGCATCTATAAATAG
ACAAGCACAATGAAAATCCTCATCATCCTTCACCACAATTCAAATATTATAGT
TGAAGCATAGTAGTAGAATCCAACAACAATCTAGAG

Seq #22 Rice cysteine peptidase promoter seq

CCAGGCTTCATCCTAACCATTACAGGCAAGATGTTGTATGAAGAAGGGCGAA
CATGCAGATTGTTAAACTGACACGTGATGGACAAGAATGACCGATTGGTGAC
CGGTCTGACAATGGTCATGTCGTCAGCAGACAGCCATCTCCCACGTCGCGC
CTGCTTCCGGTGAAAGTGGAGGTAGGTATGGGCCGTCCCGTCAGAAGGTGA
TTCGGATGGCAGCGATACAAATCTCCGTCCATTAATGAAGAGAAGTCAAGTT
GAAAGAAAGGGAGGGAGAGATGGTGCATGTGGGATCCCCTTGGGATATAAA
AGGAGGACCTTGCCCACTTAGAAAGGAGAGGAGAAAGCAATCCCAGAAGAA
TCGGGGGCTGACTGGCACTTTGTAGCTTCTTCATACGCGAATCCACCAAAAC
ACAGGAGTAGGGTATTACGCTTCTCAGCGGCCCGAACCTGTATACATCGCC
CGTGTCTTGTGTGTTTCCGCTCTTGCGAACCTTCCACAGATTGGGAGCTTAG
AACCTCACCCAGGGCCCCCGGCCGAACTGGCAAAGGGGGGCCTGCGCGGT
CTCCCGGTGAGGAGCCCCACGCTCCGTCAGTTCTAAATTACCCGATGAGAA
AGGGAGGGGGGGGGGGGAAATCTGCCTTGTTTATTTACGATCCAACGGATT
TGGTCGACACCGATGAGGTGTCTTACCAGTTACCACGAGCTAGATTATAGTA
CTAATTACTTGAGGATTCGGTTCCTAATTTTTTACCCGATCGACTTCGCCATG
GAAAATTTTTTATTCGGGGGAGAATATCCACCCTGTTTCGCTCCTAATTAAGA
TAGGAATTGTTACGATTAGCAACCTAATTCAGATCAGAATTGTTAGTTAGCGG
CGTTGGATCCCTCACCTCATCCCATCCCAATTCCCAAACCCAAACTCCTCTT
CCAGTCGCCGACCCAAACACGCATCCGCCGCCTATAAATCCCACCCGCATC
GAGCCTATCAAGCCCAAAAAACCACAAACCAAACGAAGAAGGAAAAAAAAAG
GAGGAAAAGAAAAGAGGAGGAAAGCGAAGAGGTTGGAGAGAGACGCTCGT
CTCCACGTCGCCGCC

Seq #23 Barley D-Hordian promoter

CTTCGAGTGCCCGCCGATTTGCCAGCAATGGCTAACAGACACATATTCTGCC
AAAACCCCAGAACAATAATCACTTCTCGTAGATGAAGAGAACAGACCAAGAT
ACAAACGTCCACGCTTCAGCAAACAGTACCCCAGAACTAGGATTAAGCCGAT
TACGCGGCTTTAGCAGACCGTCCAAAAAAACTGTTTTGCAAAGCTCCAATTC
CTCCTTGCTTATCCAATTTCTTTTGTGTTGGCAAACTGCACTTGTCCAACCGA

TTTTGTTCTTCCCGTGTTTCTTCTTAGGCTAACTAACACAGCCGTGCACATAG CCATGGTCCGGAATCTTCACCTCGTCCCTATAAAAGCCCAGCCAATCTCCAC AATCTCATCATCACCGAGAACACCGAGAACCACAAAACTAGAGATCAATTCA TTGACAGTCCACCG Rice Glb promoter and Gt1 leader coding sequence	
Other storage body leader sequences Bx7 #24 bx7 signal peptide seq ATGGCTAAGCGCCTGGTCCTCTTTGCGGCAGTAGTCGTCGCCCTCGTGGCT CTCACCGCC Seq #25 Glub-2 signal peptide seq ATGGCAACTACCATTTTCTCTCGTTTTTCTATATACTTTTGTGCTATGCTATTA TGCCAGGGTTCTATGGCC Seq #26 Gt3 signal peptide seq ATGTGGACATTAACAAACTCTATCTTAACATCTAGTCGATCACTACTTTACTTC ACTATAAAAGGACCAACATATATCATCCATT Seq #27 Glub-1 signal peptide seq ATGGCGAGTTCCGTTTTCTCTCGGTTTTCTATATACTTTTGTGTTCTTCTATTA TGCCATGGTTCTATGGCC Seq #28 proalmin signal peptide seq ATGAAGATCATTTTCGTATTTGCTCTCCTTGCTATTGTTGCATGCAACGCTTC TGCACGGTTTGATGCT Seq #29 Rice cysteine peptidase signal peptide seq ATGGCCGCCCGCGCCGCCGCCGCCGCGTTCCTGCTGCTGCTCATCGTCGT TGGTCACCGCGCC Seq #30 D- Hodian signal peptide ATGGCTAAGCGGCTGGTCCTCTTTGTGGCGGTAATCGTCGCCCTCGTGGCT CTCACCACCGCCOther monocot maturation specific promoter sequences	24-30
O2 transcription factor sequence ATGGAGCACGTCATCTCAATGGAGGAGATCCTCGGGCCCTTCTGGGAGCTG CTACCACCGCCAGCGCCAGAGCCAGAGCGAGAGCAGCCTCCGGTAACCGG CATCGTCGTCGGCAGTGTCATAGACGTTGCTGCTGCTGGTCATGGTGACGG GGACATGATGGATCAGCAGCACGCCACAGAGTGGACCTTTGAGAGGTTACT AGAAGAGGAGGCTCTGACGACAAGCACACCGCCGCCGGTGGTGGTGGTGC CGAACTCTTGTTGCTCAGGCGCCCTAAATGCTGACCGGCCGCCGGTGATGG AAGAGGCGGTAACTATGGCGCCTGCGGCGGTGAGTAGTGCCGTAGTAGGT GACCCCATGGAGTACAATGCCATACTGAGGAGGAAGCTGGAGGAGGACCTC GAGGCCTTCAAAATGTGGAGGGCGGCCTCCAGTGTTGTGACCTCAGATCAA CGTTCTCAAGGCTCAAACAATCACACTGGAGGTAGCAGCATCAGGAATAATC CAGTGCAGAACAAGCTGATGAACGGCGAAGATCCAATCAACAATAACCACG CTCAAACTGCAGGCCTTGGCGTGAGGCTTGCTACTAGCTCTTCCTCGAGAG ATCCTTCACCATCAGACGAAGACATGGACGGAGAAGTAGAGATTCTGGGGT TCAAGATGCCTACCGAGGAAAGAGTGAGGAAAAGAAAGGAATCCAATAGAG AATCAGCCAGACGCTCGAGATACAGGAAAGCCGCTCACCTGAAAGAACTGG AAGACCAGGTAGCACAGCTAAAAGCCGAGAATTCTTGCCTGCTGAGGCGCA TTGCCGCTCTGAACCAGAAGTACAACGACGCTAACGTCGACAACAGGGTGC	31

TGAGAGCGGACATGGAGACCCTAAGAGCTAAGGTGAAGATGGGAGAGGACT CTCTGAAGCGGGTGATAGAGATGAGCTCATCAGTGCCGTCGTCCATGCCCA TCTCGGCGCCGACCCCCAGCTCCGACGCTCCAGTGCCGCCGCCGCCTATC CGAGACAGCATCGTCGGCTACTTCTCCGCCACAGCCGCAGACGACGATGCT TCGGTCGGCAACGGTTTCTTGCGACTGCAAGCTCATCAAGAGCCTGCATCC ATGGTCGTCGGTGGAACTCTGAGCGCCACAGAGATGAACCGAGTAGCAGCA GCCACGCATTGCGCGGGGGCCATGGAGCACATCCAGACGGCGATGGGATC CATGCCGCCGACCTCCGCCTCCGGATCTACACCGCCGCCGCAGGATTATGA GCTGCTGGGTCCAAATGGGGCCATACACATGGACATGTATTAG	
PBF transcription factor sequence ATGGACATGATCTCCGGCAGCACTGCAGCAACATCAACACCCCACAACAAC CAACAGGCGGTGATGTTGTCATCCCCCATTATAAAGGAGGAAGCTAGGGAC CCAAAGCAGACACGAGCCATGCCCCAAATAGGTGGCAGTGGGGAGCGTAA GCCGAGGCCGCAACTACCTGAGGCGCTCAAGTGCCCACGCTGCGACTCCA ACAACACCAAGTTTTGCTACTACAACAATTATAGCATGTCACAACCACGCTAC TTTTGCAAGGCTTGCCGCCGCTATTGGACACATGGTGGTACCCTCCGCAAT GTCCCCATTGGTGGTGGGTGTCGCAAGAACAAACATGCCTCTAGATTTGTCT TGGGCTCTCACACCTCATCGTCCTCATCTGCTACCTATGCACCATTATCCCC TAGCACCAACGCTAGCTCTAGCAATATGAGCATCAACAAACATATGATGATG GTGCCTAACATGACGATGCCTACCCCAACGACAATGGGCTTATTCCCTAATG TGCTCCCAACACTTATGCCGACAGGTGGAGGCGGGGGCTTTGACTTCACTA TGGACAACCAACATAGATCATTGTCCTTCACACCAATGTCTCTACCTAGCCA GGGGCCAGTGCCTATGCTGGCTGCAGGAGGGAGTGAGGCAACACCGTCTT TCCTAGAGATGCTGAGAGGAGGGATTTTTCATGGTAGTAGTAGCTATAACAC AAGTCTCACGATGAGTGGTGGCAACAATGGAATGGACAAGCCATTTTCGCTG CCATCATATGGTGCAATGTGCACAAATGGGTTGAGTGGCTCAACCACTAATG ATGCCAGACAACTGGTGGGGCCTCAGCAGGATAACAAGGCCATCATGAAGA GCAGTAATAACAACAATGGTGTATCATTGTTGAACCTCTACTGGAACAAGCA CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAG GGACAATAA	32
Reb transcription factor sequence ATGGAGCGGGTGTTCTCCGTGGAGGAGATCTCCGACCCATTCTGGGTCCCG CCTCCGCCGCCGCAGTCGGCGGCGGCGGCCCAGCAGCAGGGCGGCGGCG GCGTGGCTTCGGGAGGTGGTGGTGGTGTAGCGGGGGGCGGCGGCGGCGG GAACGCGATGAACCGGTGCCCGTCGGAGTGGTACTTCCAGAAGTTTCTGGA GGAGGCGGTGCTCGATAGCCCCGTCCCGAACCCTAGCCCGAGGGCCGAAG CGGGAGGGATCAGGGGCGCAGGAGGGGTGGTGCCGGTCGATGTTAAGCA GCCGCAGCTCTCGGCGGCGGCGACGACGAGCGCGGTGGTGGACCCCGTG GAGTACAACGCGATGCTGAAGCAGAAGCTGGAGAAGGACCTCGCCGCGGT CGCCATGTGGAGGGTACAGCCATTCTCCCCCCCTCTAGTACTCGAGAGCTT ACTGAGATCGGCAATGCTAGCTACTGTTTGCATCGAATGTTTATAGGTATTTA GATCGGGCATTTCTATAGACCAATGGCGTCCATGGTCTTGCAATGCGCTCTG TTGAGTGTCGGTGGTTGGTTCGACTCATAGTATGTAGGGTTGTGCGTATGTA CAAACGGAAGCTTCATAGACCTCGGTATTGAGATTGCGATATCGATGCAACC TGCGAATTGGCGATGTAATCAGTCATATTCTTACTAAACTGCGAGACAGTGG TTTGTTTGCAATTGCAATATTTTTGTATGGGGCTGCTTAAACTGTCATTGCCT TTTTAGATTGGCAATATGTGACTTTATGCAAGTATTTGATTGGGCGGATCCAG GAACAAAAAGTTGGGGGGATTCAACATACCGAGTACACTGGCATAAACACAT CATCTCAGTATTAAACTATGCTAAAATGCTATTAAGAGACCTTTAGCACCTCT TATCTTATCAACCATGGTGAAAAAATTGAAGGGGGGACTCAGGGGGGTATCC ATGGGTCCGATGGGTGCAGGGGGGACTGAGTCCCCCCTGCACCCACGTTG AATCCGCCCTGGCATGCGTATAAGCTGTCACAGCCATTTCTAGGTGCTTGTG CTTAGTTGGGTGATGTCAGCTTAATTTGTCTTTTCTATGTCGTCATCGATTTT CTAAGAAACGAAAAATAGCCTATTTATGTGCTCCAGAATTTGATGATCCCTGG CCCTTCATTTGCTGAAATTAGCCTATTTGTTGGTTGCCCTTCAGTTTTTTCCC AGCTTATGTTGTTGCAATGTGTGGCTATGCCTCGTTTTGTGCCCTATAATTTA TTATTTGCAATTCATTTTTGTACATGACTTAAAATGACACTAGAGCAACATGCA CTGATTGGTTATCCTATAATCATTTATGTAGTTCTGTTCATTTTATCATGCTAG	33

CTCATGTCATTTTCATCTTCAGGCCTCTGGCACAGTTCCACCTGAGCGTCCT
GGAGCTGGTTCATCCTTGCTGAATGCAGATGTTTCACACATAGGCGCTCCTA
ATTCCATCGGAGGTACTTATCTTATCTGGTTACATTTTCAGATTGTTATGAAA
CTACCCAAATATCCTGCACAATTGCATGGGATTAAATTTTAGTTTCTTTGAAAT
AGAAGTAGAGTTGTATTGCTGTCACGTCATCAAATAGTTCTGAAGCTATGAAT
AAATAAGTTCCGCATTTGTTAGTGATTCTTTGAACATTAGAATTGTTATGCTTA
AGTAGATAGGGTTATGTTTGTTTGGAGTTCCCTTAAATCATTTCATTGCTGAC
TGCCAGCTGGCAGGAGCATTTGTTGTTGCCTTGACCATGAATGAAGACCTTC
CTGTTCTGAGTGCTCACAAGAAAACATATTTTGATTAATGCACCTTGAATCCT
TAGGATCTTGCAAAGATGGGCACTTAGCTTTAGAATTGAGTAGTACTTAAATA
GCTGTTGTTATCATGATTTGTCCTGTAGTGAAATGTCGACAAAACAGGAATG
CTACTTTTGACTTCTGATATTTCATGCCTGGCTTTACTTATGCTCTGTTTGGAA
CATGGGCACATATCAGGCAATGCTACTCCAGTTCAAAACATGCTAAGTGGCC
CAAGTGGGGGATCGGGCTCACAGTTGGTACAGAATGTTGATGTCCTTGTAAA
GCAGCCCACCAGCTCTTCATCAAGGGAGCAGTCAGATGATGATGACATGAA
GGGAGAAGCTGAGACCACTGGAACTGCAAGACCTGCTGATCAAAGATTACA
ACGAAGGTGATCATTCATTGCTTCCTTGTAATATAGATTCTGTACATAATTAA
CCTACCTCGTCATGCATGCATGTGTCCTATTTTCACCTTAGCCCTTTCAGTTG
GATTTCCACTTTCATCCGGTAGCCTTTCAGTTTCCTATTGCATCGCATATATG
ATCTTTTACCTACCATATTAGTTCTCTGTGTGCCATACTCAGTGCTTAGTGTC
TCGAGCAAGAGAGGAATTTGTATGGCTATTACACGTAGCACTTTGCTCTCTA
CTTGTTTATTGACATAAGCAATTTGGGATGAATTAAATCTGAGTTCACATCAT
ATTCCTTATGTCACAAGTTTCTGAAACCGATTGTATCTAGTATCTGGTTGATG
CACCCCCATCTTGGATTTGCAAATCAAAGTTATACTCCCTAGAGAGCTTTACC
TTTCATAAAGCAATTACCCCAATAAACCACGGATTTGATAGCTATTGACTATG
ATTACCAGAATTCATTTGGCAGCTATTTTCTCAATTTAAGTTTGGTATTAGTCT
CAGTTGGCTGTAAAATAATGTCACGGTAGGGTACATGTATGTGCAGCATACA
AGGTATGGGTGAGTTATGATATGGACAGTGTGTACACCCCACATTTGCTCAC
TAAAATCAAAATATTCAAACGTCACGTGATGATATGGTGGATTGCATTATACC
TTGTATTGTTTATTATGTTACTTGTGCTAGACAATAATATAGGCTGTTCTTTTG
GGTGATTTTGTATGAAGATGTTGAGCAAGCACTTCTCGATATAATGCTAGTTT
TGTTGACCTGTTCCAGGAAGCAATCCAATCGGGAGTCAGCCAGGCGCTCAA
GAAGCAGAAAGGCAGCTCACTTGAATGAGCTGGAGGCACAGGTGTGATAGT
TCACATAGTTATTTTCGATAAGACATAAAATCCTAAATTACTGGCTACTGACTT
CAGTTATGGATTTACTTGTTACAGGTATCGCAATTAAGAGTCGAGAACTCCTC
GCTGTTAAGGCGTCTTGCTGATGTTAACCAGAAGTACAATGATGCTGCTGTT
GACAATAGAGTGCTAAAAGCAGATGTTGAGACCTTGAGAGCAAAGGTATGCT
ATATATGCCTTTTGCAATATGCATCCCATGGATTGCTACTTTGGCTTGTTTCA
AACTTTCAACGTGACTTGTGTACCCTGTTATTAGAAGAATAATCCCGCCTACC
ATTATACTCTATAAATCACCATTTGGCCAGTCCAAACATGATTATTAAATCAG
GTCAATCTGAACATTGAAATGTATCAAAAATTCGCAGGTGAAGATGGCAGAG
GACTCGGTGAAGCGGGTGACAGGCATGAACGCGTTGTTTCCCGCCGCTTCT
GATATGTCATCCCTCAGCATGCCATTCAACAGCTCCCCATCTGAAGCAACGT
CAGACGCTGCTGTTCCCATCCAAGATGACCCGAACAATTACTTCGCTACTAA
CAACGACATCGGAGGTAACAACAACTACATGCCCGACATACCTTCTTCGGCT
CAGGAGGACGAGGACTTCGTCAATGGCGCTCTGGCTGCCGGCAAGATTGG
CCGGCCAGCCTCGCTGCAGCGGGTGGCGAGCCTGGAGCATCTCCAGAAGA
GGATGTGCGGTGGGCCGGCTTCGTCTGGGTCGACGTCCTGA

WHAT IS CLAIMED IS:

1. A method of producing a heterologous polypeptide in a grain of a plant, comprising the steps of:

(a) culturing a transformed plant to form a grain-producing transformed plant; and

(b) recovering transgenic grains containing the heterologous polypeptide from the grain-producing transformed plant;

wherein the transformed plant comprises a first chimeric gene;

wherein the first chimeric gene comprises (i) a rice glutelin Gt1 promoter; (ii) a nucleic acid leader encoding a rice glutelin Gt1 signal peptide; and (iii) a heterologous nucleic acid that encodes the heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide.

2. A method of producing a heterologous polypeptide in a grain of a plant, comprising the steps of:

(a) culturing a transformed plant to form a grain-producing transformed plant; and

(b) recovering transgenic grains containing the heterologous polypeptide from the grain-producing transformed plant;

wherein the transformed plant comprises a first chimeric gene;

wherein the first chimeric gene comprises (i) a rice globulin Glb promoter; (ii) a nucleic acid leader encoding a rice glutelin Gt1 signal peptide; and (iii) a heterologous nucleic acid that encodes the heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide.

3. The method of claim 1 or 2, wherein the method further comprises the steps of transforming a plurality of plant cells; culturing a plurality of transformed plant cells to form a plurality of grain-producing transformed plants; and selecting for further culturing a transformed plant that exhibits high level expression of the heterologous polypeptide in its grains.

4. The method of claim 1 or 2, wherein the transformed plant that is cultured to form a grain-producing transformed plant is generated from a grain of a parent plant that has been transformed with the first chimeric gene.

5. The method of claim 1 or 2, wherein the transformed plant further comprises a transcription factor-encoding nucleic acid that encodes at least one heterologous transcription factor that is capable of enhancing expression of the heterologous polypeptide.

6. The method of claim 5, wherein the heterologous transcription factor is selected from the group consisting of O2 and PBF and an active fragment thereof.

7. The method of claim 2, wherein the transformed plant further comprises a transcription factor-encoding nucleic acid that encodes at least one heterologous transcription factor selected from the group consisting of Reb, O2 and PBF, and an active fragment thereof.

8. The method of claim 5, wherein the transformed plant further comprises one or more transcription factor-encoding nucleic acids that encode two heterologous transcription factors that are capable of enhancing expression of the heterologous polypeptide.

9. The method of claim 8, wherein the two heterologous transcription factors are O2 and PBF.

10. The method of claim 1 or 2, wherein the heterologous nucleic acid that encodes the heterologous polypeptide is codon optimized for expression in a dicot or a monocot.

11. The method of claim 1 or 2, wherein the plant is a dicot.

12. The method of claim 1 or 2, wherein the plant is a monocot.

13. The method of claim 12, wherein the monocot is rice, wheat or barley.

14. The method of claim 1 or 2, wherein the heterologous polypeptide is a non-plant storage polypeptide.

15. The method of claim 1 or 2, wherein the heterologous polypeptide is selected from the group consisting of: antibodies, cytokines, lymphokines, chemokines, hormones, growth factors, coagulation factors, anti-infectives, and cytotoxins.

16. The method of claim 1 or 2, wherein the heterologous polypeptide is a human or non-human animal polypeptide.

17. The method of claim 1 or 2, wherein the heterologous polypeptide is a milk polypeptide.

18. The method of claim 1 or 2, wherein the heterologous polypeptide is other than a milk polypeptide.

19. The method of claim 1 or 2, wherein the heterologous polypeptide is an anti-inflammatory molecule.

20. The method of claim 1 or 2, wherein the heterologous polypeptide is intestinal trefoil factor (ITF) or an active fragment thereof.

21. The method of claim 1 or 2, wherein the heterologous polypeptide is IgA or an active fragment thereof.

22. The method of claim 1 or 2, wherein the heterologous polypeptide is selected from the group consisting of: lactoferrin, lysozyme, lactoferricin, intestinal trefoil factor (ITF), epidermal growth factor (EGF), keratinocyte growth factor (KGF), insulin-like growth factor I (IGF-I), lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin, immunoglobulins, alpha-lactalbumin, beta-lactoglobulin, alpha-casein, beta-casein, albumin, fibrinogen, and a protease inhibitor.

23. The method of claim 1, wherein the first chimeric gene is selected from the group consisting of:

(a) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized lysozyme coding sequence;

(b) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized lactoferrin coding sequence;

(c) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized EGF;

(d) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized IGF-I;

(e) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized haptocorrin;

(f) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and codon optimized IGF-I;

(g) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized lactahedrin;

(h) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized lactoperoxidase; and

(i) a chimeric gene comprising a rice glutelin Gt1 promoter, a rice glutelin Gt1 signal peptide, and a codon optimized α_1-antitrypsin.

24. The method of claim 2, wherein the first chimeric gene is selected from the group consisting of: (a) a rice globulin Glb promoter, a rice glutelin Gt1 signal peptide, and codon optimized haptocorrin; and (b) a rice globulin Glb promoter, a rice glutelin Gt1 signal peptide, and codon optimized lysozyme.

25. The method of claim 1, 2, 5 or 8, wherein the transformed plant comprises a second chimeric gene, wherein the second chimeric gene comprises: (i) an aleurone-specific promoter; and (ii) the heterologous nucleic acid encoding the heterologous polypeptide, operably linked together to enable expression of the heterologous polypeptide.

26. An expression vector comprising a first chimeric gene that comprises: (i) a rice glutelin Gt1 promoter; (ii) a nucleic acid leader encoding a rice glutelin Gt1 signal peptide; and (iii) a heterologous nucleic acid that encodes a heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide.

27. An expression vector comprising a first chimeric gene that comprises: (i) a rice globulin Glb promoter; (ii) a nucleic acid leader encoding a rice glutelin Gt1 signal peptide; and (iii) a heterologous nucleic acid that encodes a heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide.

28. The expression vector of claim 26 or 27, further comprising a selectable marker.

29. The expression vector of claim 28, wherein the selectable marker is an antibiotic resistance gene.

30. The expression vector of claim 28, wherein the heterologous polypeptide is a non-plant storage polypeptide.

31. The expression vector of claim 28, wherein the heterologous polypeptide is selected from the group consisting of: antibodies, cytokines, lymphokines, chemokines, hormones, growth factors, coagulation factors, anti-infectives, and cytotoxins.

32. The expression vector of claim 28, wherein the heterologous polypeptide is a human or non-human animal polypeptide.

33. The expression vector of claim 28, wherein the heterologous polypeptide is a milk polypeptide.

34. The expression vector of claim 28, wherein the heterologous polypeptide is other than a milk polypeptide.

35. The expression vector of claim 28, wherein the heterologous polypeptide is selected from the group consisting of: lactoferrin, lysozyme, lactoferricin, intestinal trefoil factor (ITF), epidermal growth factor (EGF), keratinocyte growth factor (KGF), insulin-like growth factor I (IGF-I), lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin, immunoglobulins, immunoglobulins, alpha-lactalbumen, beta-lactoglobulin, alpha-casein, beta-casein, albumin, and fibrinogen.

36. The expression vector of claim 28, further comprising a second chimeric gene that comprises: (i) an aleurone-specific promoter; and (ii) the heterologous nucleic acid that encodes the heterologous polypeptide, operably linked to enable expression of the heterologous polypeptide.

37. The expression vector of claim 28 or 36, further comprising at least one transcription factor-encoding nucleic acid that encodes at least one heterologous transcription factor that is capable of enhancing expression of the heterologous polypeptide.

38. A transformed host comprising the expression vector of any of claims 26, 27, or 36.

39. The transformed host of claim 38, wherein the host is a plant cell, a plant or a grain.

40. The transformed host of claim 38, further comprising at least a first transcription factor-encoding nucleic acid that encodes at least a first heterologous transcription factor that is capable of enhancing expression of the heterologous polypeptide in the transformed host.

41. The transformed host of claim 40, further comprising a second transcription factor-encoding nucleic acid that encodes a second heterologous transcription factor that is capable of enhancing expression of the heterologous polypeptide.

42. The transformed host of any of claims 38, 40 or 41, wherein the transformed host is a dicot.

43. The transformed host of any of claims 38, 40 or 41, wherein the transformed host is a monocot.

44. A method of making a transformed host cell comprising transfecting a host cell with the expression vector of any of claims 26, 27, 36 or 37.

45. The method of claim 44, further comprising the steps of selecting a transformed cell, and culturing the transformed cell to produce a transformed plant.

46. The method of claim 45, further comprising the step of allowing the transformed plant to produce transformed grains.

47. Transformed grains comprising a first chimeric gene that comprises (i) a rice glutelin Gt1 promoter; (ii) a nucleic acid leader encoding a rice glutelin Gt1 signal peptide; and (iii) a heterologous nucleic acid that encodes a heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide.

48. Transformed grains comprising a first chimeric gene that comprises (i) a rice globulin Glb promoter; (ii) a nucleic acid leader encoding a rice glutelin Gt1 signal peptide; and (iii) a heterologous nucleic acid that encodes a heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide.

49. The transformed grains of claim 47 or 48, further comprising a second chimeric gene that comprises (i) an aleurone-specific promoter; and (ii) the heterologous nucleic acid that encodes the heterologous polypeptide, operably linked to enable expression of the heterologous polypeptide.

50. The transformed grains of any of claims 47, 48 or 49, further comprising at least one transcription factor-encoding nucleic acid that encodes at least one heterologous transcription factor that is capable of enhancing expression of the heterologous polypeptide.

51. The transformed grains of any of claims 47, 48, 49 or 50, wherein the transformed grains have been wet or dry milled.

52. The transformed grains of any of claims 47, 48, 49 or 50, wherein the transformed grains have been subjected to an extraction process.

53. A heterologous polypeptide produced by the process of claim 1 or 2.

54. A codon optimized nucleic acid molecule for expression of polypeptides in monocot, wherein the nucleic acid is selected from the group consisting of:

(i) a h LF nucleic acid encoding human lactoferrin;

(ii) a hLZ nucleic acid encoding human lysozyme;

(iii) a hEGF nucleic acid encoding human epidermal growth factor;

(iv) hIGF-I nucleic acid encoding human insulin-like growth factor I;

55. The codon optimized nucleic acid molecule of claim 52, wherein nucleic acid molecule comprises a sequence selected from the group consisting of: (i) hLF nucleic acid sequence as shown in SEQ ID NO:3;

(ii) hLZ nucleic acid sequence as shown in SEQ ID NO:1;

(iii) hEGF nucleic acid sequence as shown in SEQ ID NO: 8.

(iv) hIGF-I nucleic acid sequence as shown in SEQ ID NO: 9

56. A method of producing a heterologous polypeptide in a grain of a plant, comprising the steps of:

(a) culturing a transformed plant to form a grain-producing transformed plant; and

(b) recovering grains containing the heterologous polypeptide from the grain-producing transformed plant;

wherein the transformed plant comprises a first chimeric gene and at least one heterologous transcription factor that is capable of enhancing the expression of the first chimeric gene;

wherein the first chimeric gene comprises: (i) a promoter of a storage protein gene; (ii) a nucleic acid leader encoding a signal peptide of a storage protein; and (iii) a heterologous nucleic acid that encodes the heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide.

57. The method of claim 56, wherein the transformed plant further comprises a second chimeric gene that comprises: (i) an aleurone specific promoter; and (ii) the heterologous nucleic acid, all operably linked to enable expression of the heterologous polypeptide.

58. The method of claim 56, wherein the storage protein gene is selected from the group consisting of: rice glutelins, rice oryzins, rice prolamines, barley hordeins, wheat gliadins, wheat glutenins, maize zeins, maize glutelins, oat glutelins, sorghum kafirins, millet pennisetins, and rye secalins.

59. The method of claim 56, wherein the transformed plant comprises two heterologous transcription factors that are capable of enhancing expression of the heterologous polypeptide.

60. The method of claim 59, wherein the heterologous transcription factor is selected from the group consisting of Reb, O2, and PBF.

61. The method of claim 56, wherein the signal peptide is a rice Gt1 signal peptide or rice Glb signal peptide.

62. The method of claim 56, wherein the promoter is a rice Gt1 promoter, a rice Glb promoter or a wheat glutenin promoter.

63. The method of claim 57, wherein the aleurone specific promoter a promoter of the lipid transfer protein gene Ltp1 or a promoter of chitinase gene Chi26.

64. The method of claim 56, wherein the transformed plant is a dicot.

65. The method of claim 56, wherein the transformed plant is a monocot.

66. The method of claim 65, wherein the monocot is a cereal.

67. The method of claim 66, wherein the cereal is rice, wheat, or barley.

68. The method of claim 56, wherein the heterologous polypeptide is selected from the group consisting of: antibodies, cytokines, lymphokines, chemokines, hormones, growth factors, coagulation factors, anti-infectives, and cytotoxins.

69. The method of claim 56, wherein the heterologous polypeptide is selected from the group consisting of: lactoferrin, lysozyme, lactoferricin, intestinal trefoil factor (ITF), epidermal growth factor (EGF), keratinocyte growth factor (KGF), insulin-like growth factor I (IGF-I), lactohedrin, kappa-casein, haptocorrin, lactoperoxidase, alpha-1-antitrypsin, immunoglobulins, alpha-lactalbumin, beta-lactoglobulin, alpha-casein, beta-casein, albumin, and fibrinogen.

70. An expression vector comprising a first chimeric gene that comprises (i) an aleurone-specific promoter and (ii) a heterologous nucleic acid encoding a heterologous polypeptide, operably linked to enable expression of the heterologous polypeptide.

71. The expression vector of claim 70, further comprising a transcription factor-encoding nucleic acid that encodes at least one heterologous transcription factor that is capable of enhancing expression of the heterologous polypeptide.

72. The expression vector of claim 70 or 71, further comprising a second chimeric gene that comprises: (i) a promoter of a storage protein gene; (ii) a nucleic acid leader encoding a signal peptide of a storage protein; and (iii) the heterologous nucleic acid encoding the heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide.

73. A transformed host comprising a first chimeric gene comprising an aleurone-specific promoter operably linked to a first heterologous polypeptide to enable expression of the heterologous polypeptide.

74. The transformed host of claim 73, further comprising at least one heterologous transcription factor.

75. The transformed host of claim 73 or 74, further comprising a second chimeric gene comprising (i) a promoter of a storage protein gene, (ii) a nucleic acid leader encoding a signal peptide of a storage protein, and (iii) a heterologous polypeptide, all operably linked to enable expression of the heterologous polypeptide.

76. The transformed host of claim 75, further comprising two heterologous transcription factors.

77. A composition comprising a heterologous polypeptide in transformed grain, wherein the transformed grain comprises the expression vector of claim 70, 71 or 72.

78. The composition of claim 77, wherein the transformed grain has been wet or dry milled.

79. The composition of claim 77, wherein the transformed grain has been subjected to an extraction process.

80. A method for processing a transgenic seed that comprises a heterologous polypeptide, comprising the steps of:

(a) powderizing the transgenic seed in a buffer to form bufferized powdered seed;

(b) removing cell debris from the bufferized powdered seed to obtain a supernatant; and

(c) reducing liquid from the supernatant to produce an extract that comprises the heterologous polypeptide.

81. The method of claim 80, wherein the buffer is selected from the group consisting of: saline, phosphate buffered saline (PBS), ammonium bicarbonate buffer, ammonium acetate buffer, and Tris, with or without iron supplement.

82. The method of claim 80, wherein the step of removing cell debris is performed by centrifugation, filtration or sedimentation.

83. The method of 80, wherein the step of reducing liquid from the supernatant is performed by lyophilization or spray drying.

84. The method of claim 80, wherein the heterologous polypeptide is active after processing.

85. The method of claim 80, wherein the seed is a cereal grain.

86. The method of claim 80, wherein the seed is a rice, wheat or barley seed.

87. An extract comprising a plant-produced heterologous polypeptide produced by the process of claim 80.

88. The extract of claim 87, wherein the heterologous polypeptide is selected from the group consisting of: antibodies, cytokines, lymphokines, chemokines, hormones, growth factors, coagulation factors, anti-infectives, cytotoxins and active fragments thereof.

89. The extract of claim 80, wherein the heterologous polypeptide is a milk polypeptide.

90. A method of processing transgenic grains containing a heterologous polypeptide to produce a malt syrup comprising the steps of:

(a) providing transgenic grains that contain starch and heterologous polypeptides;

(b) providing an enzymatic composition for conversion of malt to wort;

(c) combining the transgenic grains with the enzymatic composition under conditions that allow for conversion of starch in the transgenic grains to sugars;

(d) allowing the starch in the transgenic grains to be at least partially converted; and

(e) separating resulting transgenic malt syrup from resulting transgenic grain residue, each containing the heterologous polypeptide.

91. The method of claim 90, wherein the enzymatic composition comprises amylases.

92. The method of claim 90, wherein the enzymatic composition comprises a liquid malt extract comprising barley.

93. A method of processing transgenic grains containing a heterologous polypeptide to produce a malt extract comprising the steps of:

(a) providing transgenic grains that contain starch and heterologous polypeptides;

(b) allowing the transgenic grains to undergo a malting process to form malted transgenic grains; and

(c) crushing the malted transgenic grains.

94. The method of claim 93, further comprising the step of mixing the crushed and malted transgenic grains with water to obtain a liquid transgenic malt extract.

95. The method of claim 90 or 93, wherein the transgenic grains are cereal grains.

96. The method of claim 95, wherein the cereal grains are selected from the group consisting of corn, sorghum, rice, barley, rye, wheat, oats and triticale.

97. The method of claim 90 or 93, wherein the transgenic grains are grains of a dicot.

98. A transgenic malt syrup comprising a heterologous polypeptide produced in transgenic grains in a malt syrup.

99. A transgenic malt extract comprising a heterologous polypeptide produced in transgenic grains in a malt extract.

100. An edible food comprising a heterologous polypeptide in a transgenic malt syrup, a transgenic malt extract, or a transgenic grain residue in a food product.

101. A protein-containing product comprising:

a monocot seed fluor, extract, or malt composition containing monocot seed or malted-seed components, and one or more seed-produced, non-seed proteins in substantially unpurified form, and

a vehicle containing the extract in a form suitable for human or animal use.

102. The product of claim 101, wherein said vehicle is selected from the group consisting of a capsule, binder components effective to tabletize the composition, a consumable liquid, and a consumable suspension.

103. The product of claim 101, wherein the vehicle is a processed food in which the extract is mixed.

104. The product of claim 101 or 102, wherein the seed-produced, non-seed proteins are lysozyme and lactoferrin.

105. The product of claim 101, wherein the protein is a blood-clotting factor and the vehicle is a surgical dressing, powder or cream.

106. The product of claim 101, wherein the protein is an industrial enzyme, the extract is in a liquid or powder form, and the vehicle is a liquid or powder dispenser, respectively.

107. The product of claim 101, wherein the protein is a detergent enzyme and the vehicle is a detergent.

108. The product of claim 101, wherein the protein is an anti-infective, and the vehicle is selected from the group consisting of: toothpaste, mouthwash, hand soap, dish soap, dishwasher detergent, tile cleaner, bathroom cleaner, toilet cleaner.

109. The product of claim 101, wherein the weight ratio of protein: seed components is less than 5%.

110. The product of claim 101, wherein the weight ratio of protein: seed components is less than 0.5%.

111. A method for producing a selected protein other than a monocot-plant seed-storage protein, comprising the steps of:

(a) obtaining a monocot plant that has been stably transformed with a first chimeric gene having (i) a rice glutelin Gt1 promoter or a rice globulin Glb promoter, (ii) operably linked to the promoter, a leader DNA encoding a glutelin Gt1 leader sequence, and (iii) a protein-coding sequence encoding the selected protein, and linked in translation reading frame with the leader sequence, such that the leader and the protein-encoding sequences encode a fusion protein composed of the selected protein and an N-terminal glutelin Gt1 leader sequence,

(b) cultivating the plant under seed-maturation conditions,

(c) harvesting mature seeds from the cultivated plants, and

(d) extracting from the harvested seeds, an extract containing the selected protein.

112. The method of claim 111, wherein the mature seeds contain at least about 0.25% weight percent of the selected protein.

113. The method of claim 111, wherein the mature seeds contain at least about 0.5% weight percent of the mature seed weight.

114. The method of claim 111, wherein the rice glutelin Gt1 promoter and the leader DNA encoding a glutelin Gt1 leader have the sequence identified by SEQ ID NO: 15.

115. The method of claim 111, wherein the rice globulin Glb promoter and the leader DNA encoding a glutelin Gt1 leader have the sequence identified by SEQ ID NO.16

116. The method of claim 111, wherein the step of extracting comprises crushing the seeds and suspending the crushed seeds in an aqueous medium buffered to physiological pH.

117. The method of claim 111, wherein the step of extracting comprises milling mature seeds to produce a seed grain and malting the resulting seed grain.

118. A plant stably transformed with a chimeric gene comprising (i) a rice glutelin Gt1 promoter or a rice globulin Glb promoter, (ii) operably linked to said promoter, a leader DNA encoding a glutelin Gt1 leader sequence, and (iii) a protein-coding sequence encoding the selected protein, and linked in translation reading frame with the leader sequence, such that the leader and protein-encoding sequences encode a fusion protein composed of the selected protein and an N-terminal glutelin Gt1 leader sequence.

119. The plant of claim 118, wherein the plant is further transformed with one or more nucleic acid that encode at least one transcription factor that are capable of enhancing expression of the selected protein.

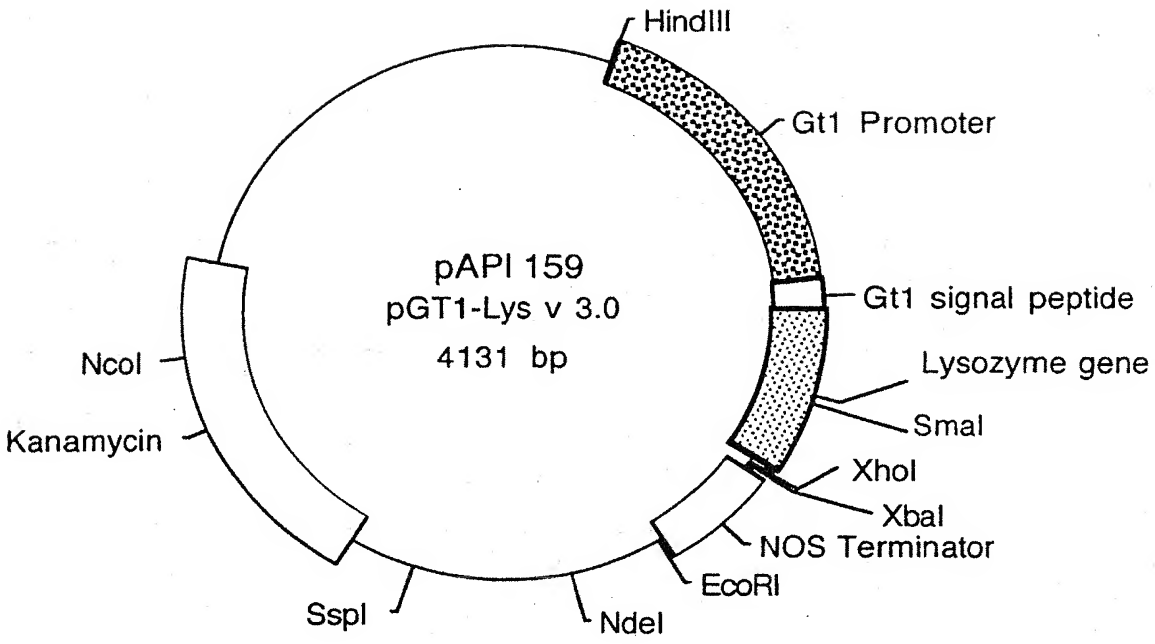

HindIII
Gt1 Promoter
pAPI 159
pGT1-Lys v 3.0
4131 bp
Gt1 signal peptide
Lysozyme gene
SmaI
XhoI
XbaI
NOS Terminator
EcoRI
NdeI
SspI
NcoI
Kanamycin

Fig. 1

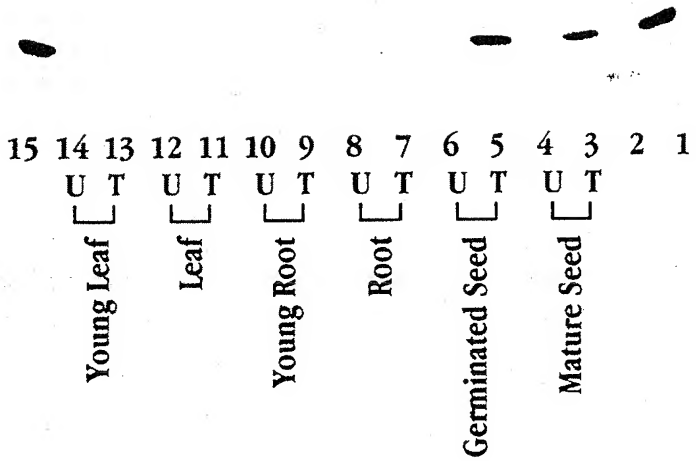

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
U T U T U T U T U T U T U T
Young Leaf
Leaf
Young Root
Root
Germinated Seed
Mature Seed

Fig. 2

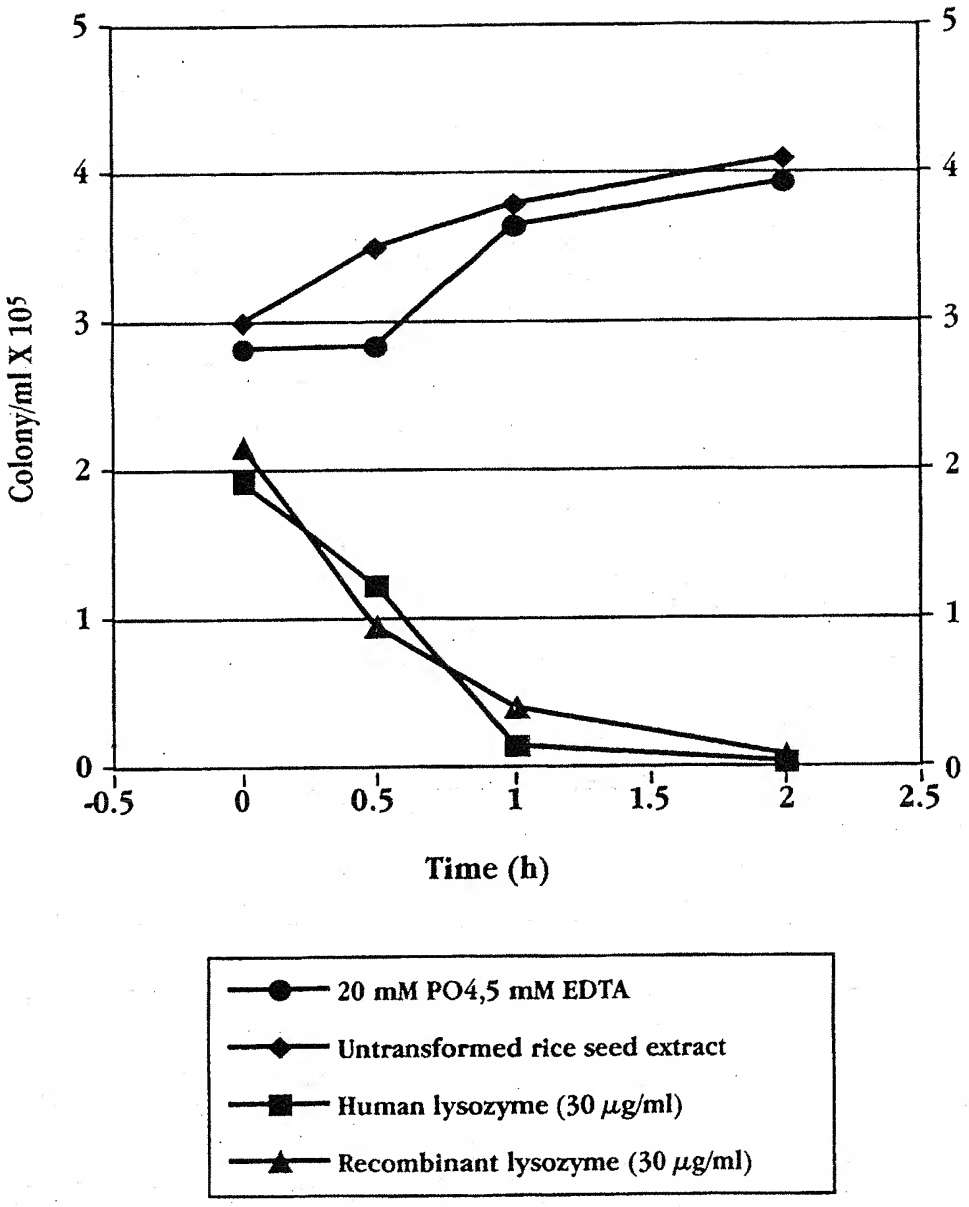

Colony/ml X 10⁵
5
4
3
2
1
0
-0.5
0
0.5
1
1.5
2
2.5
Time (h)
20 mM PO4,5 mM EDTA
Untransformed rice seed extract
Human lysozyme (30 μg/ml)
Recombinant lysozyme (30 μg/ml)

Fig. 3

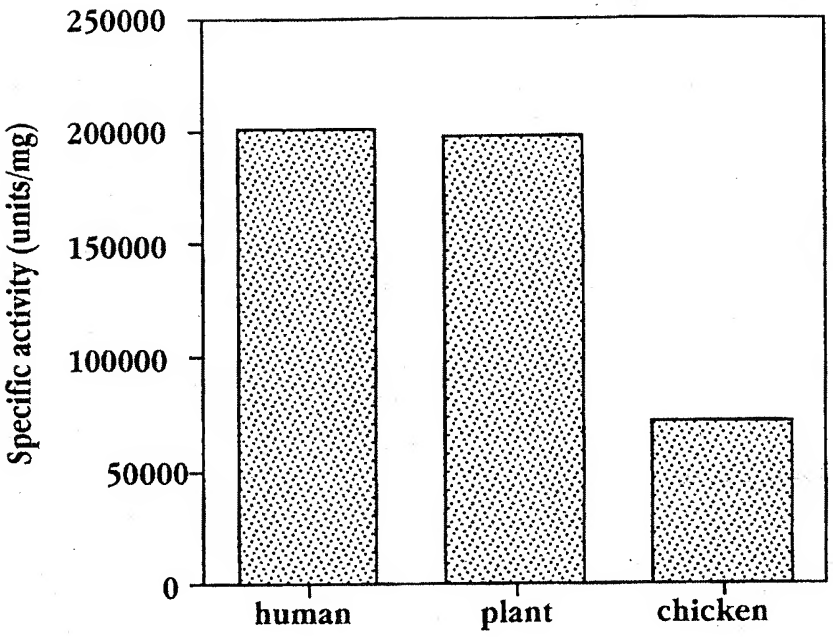

250000
200000
150000
100000
50000
0
Specific activity (units/mg)
human
plant
chicken

Fig. 4

Fig. 5A Fig. 5B

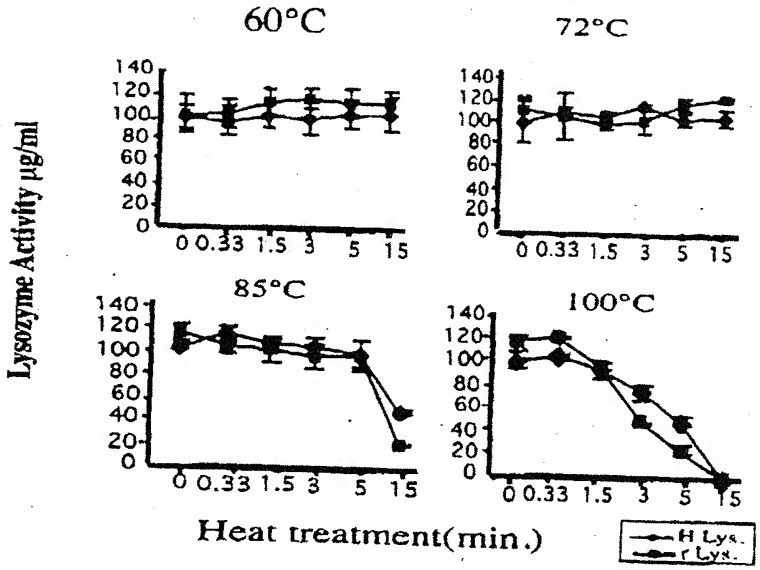


Fig. 5C Fig. 5D

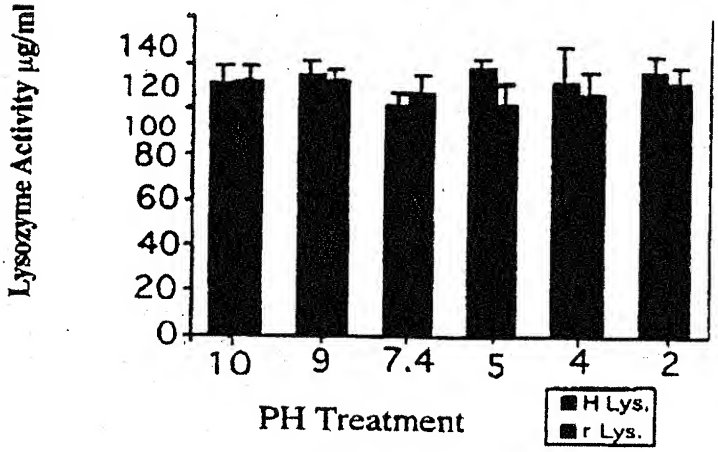


Fig. 5E

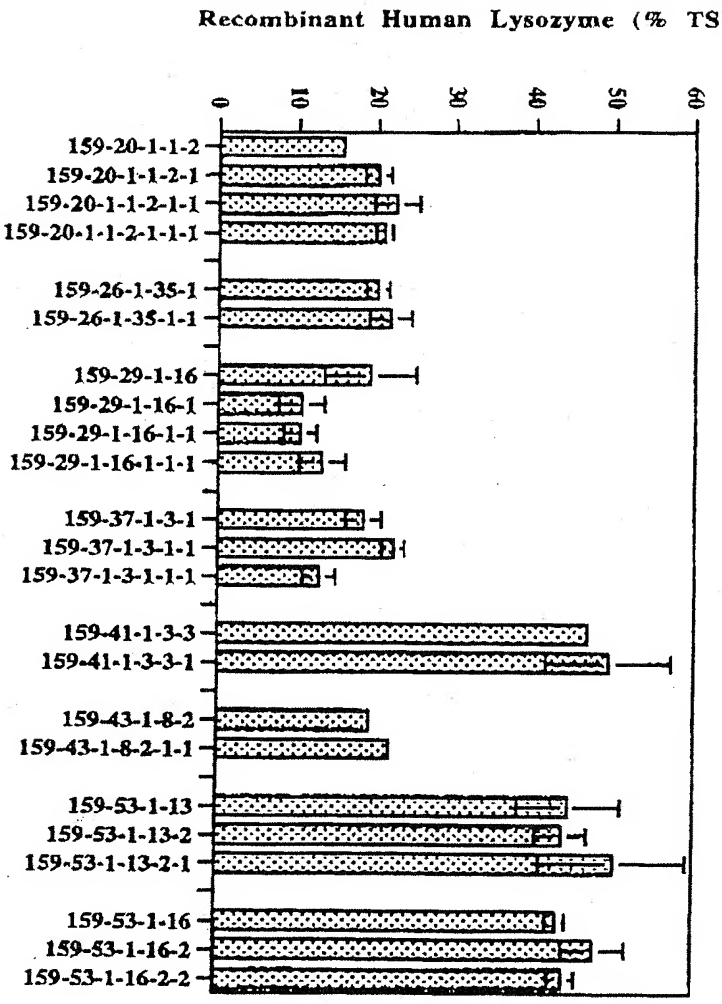


Fig. 6

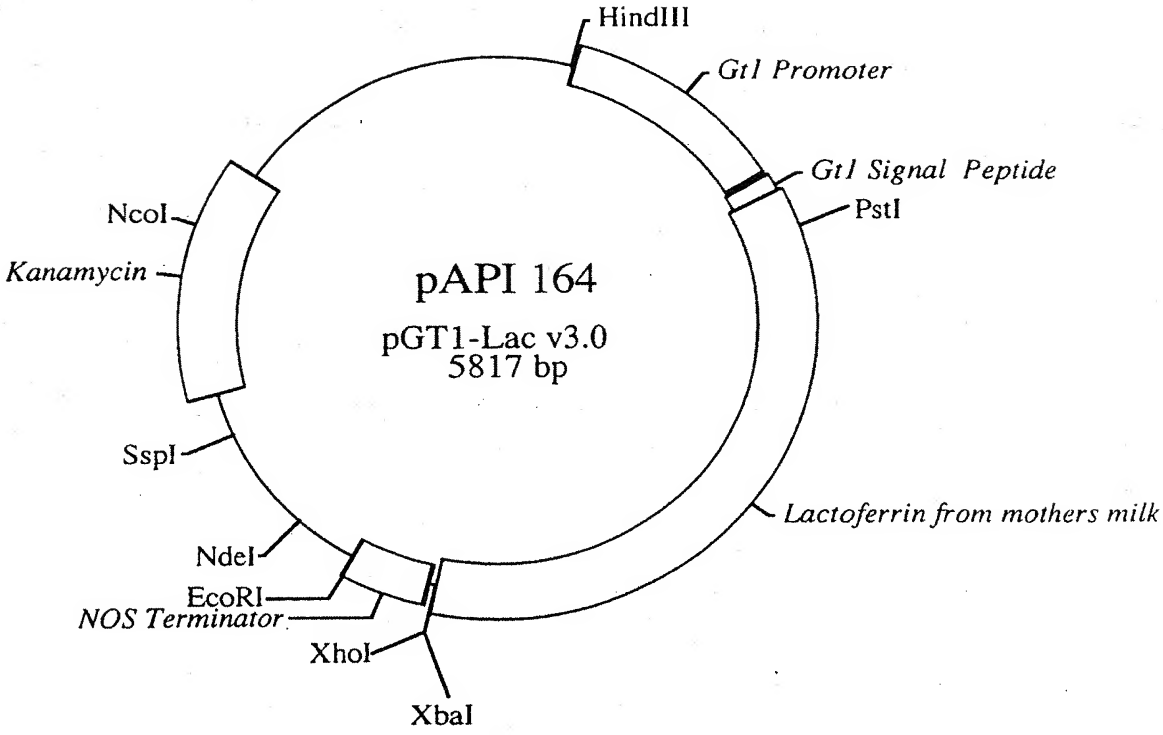


Fig. 7

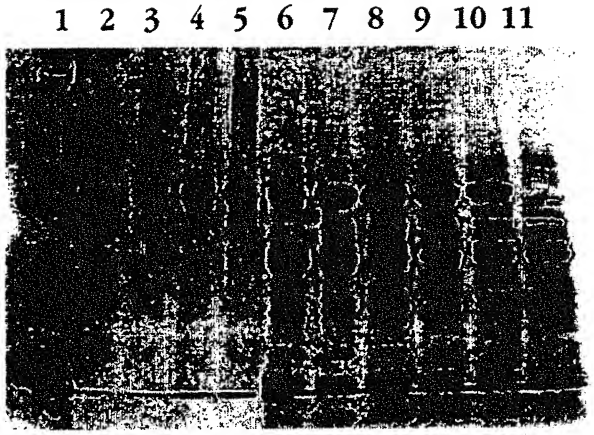


Fig. 8

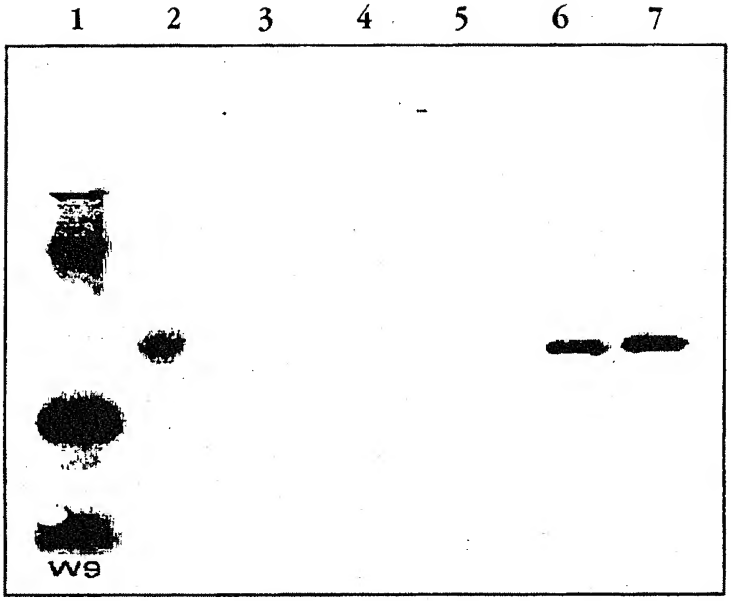


Fig. 9

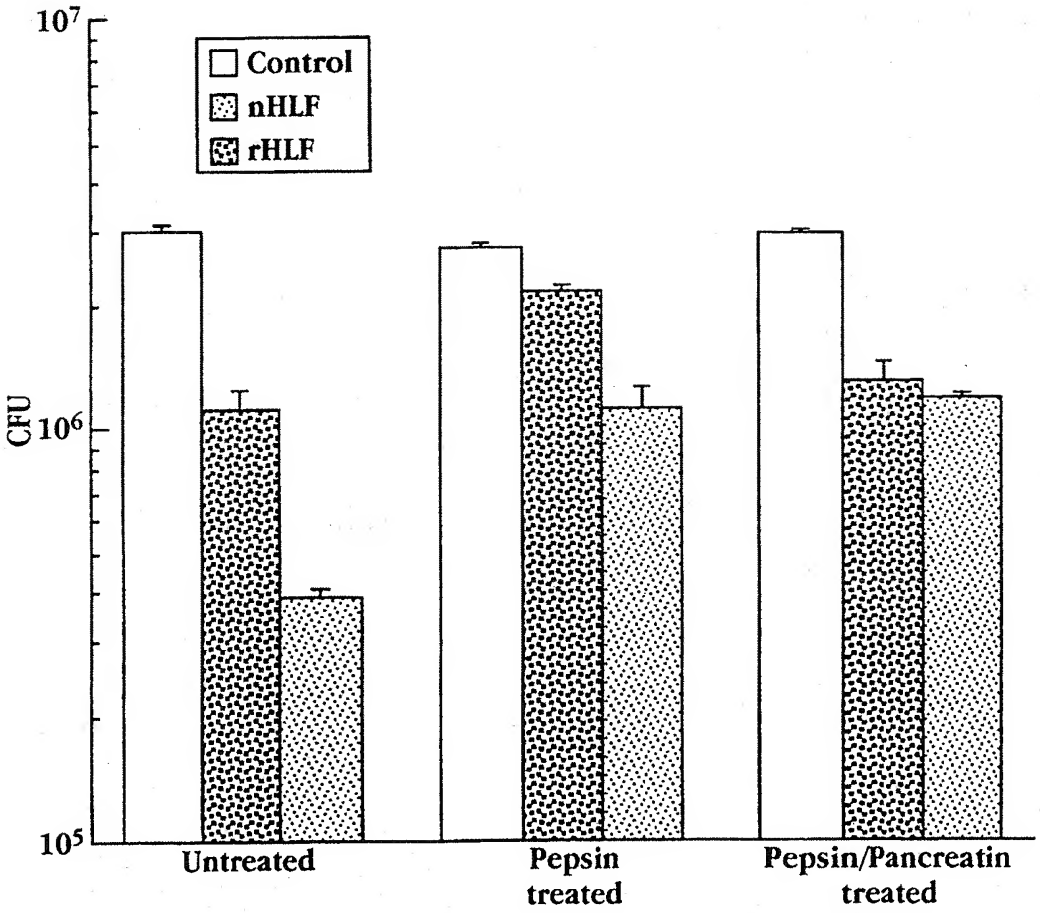


Fig. 10

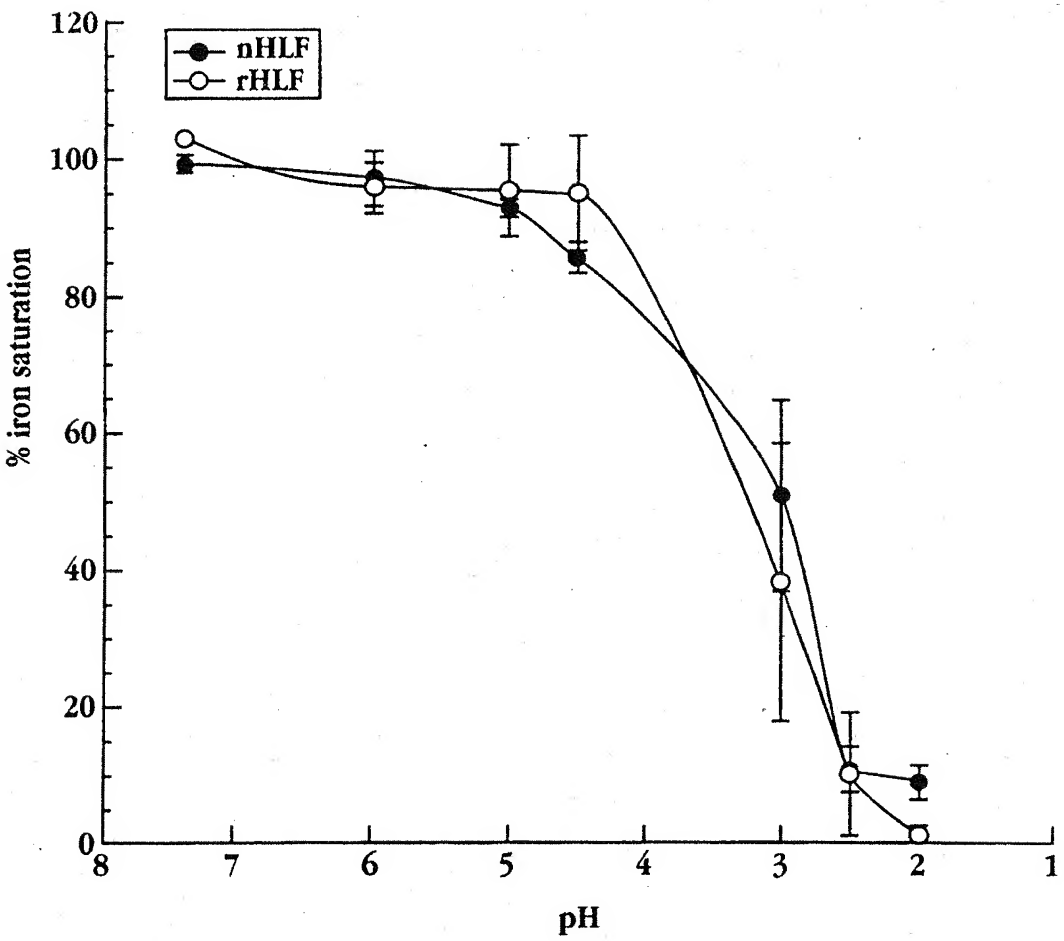


Fig. 11

Fig. 12A

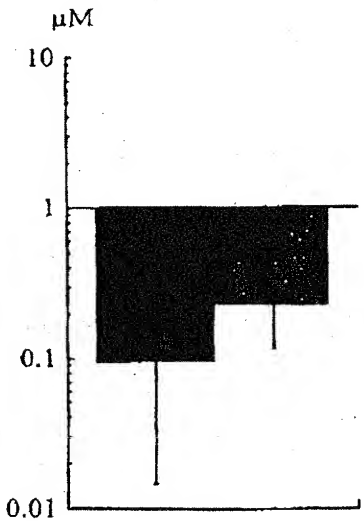


Fig. 12B

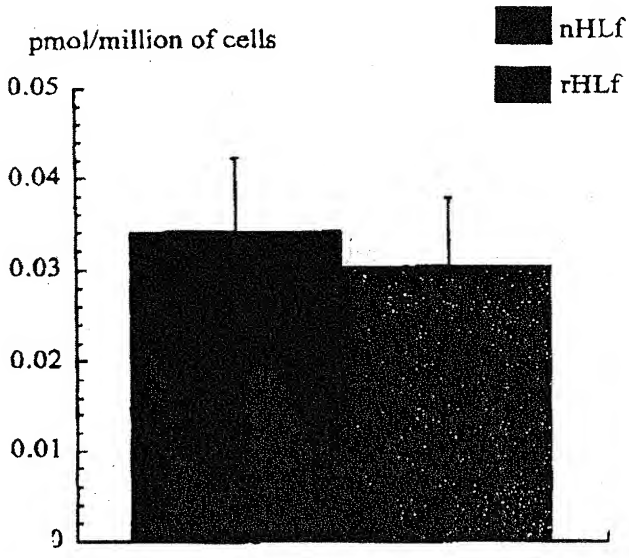


Fig. 12C

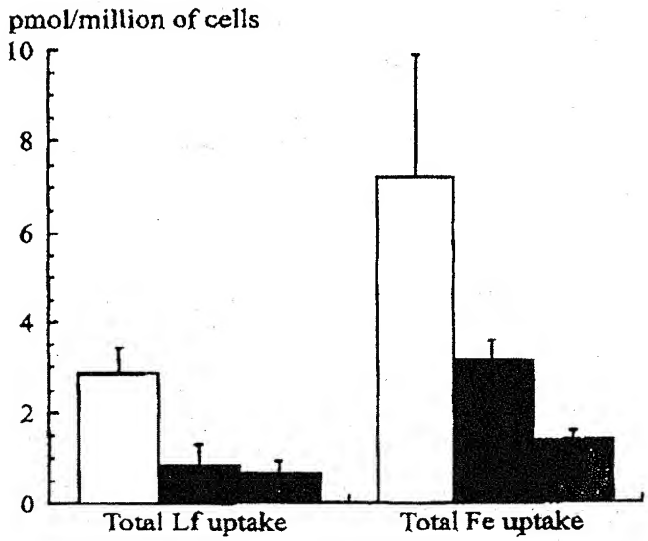
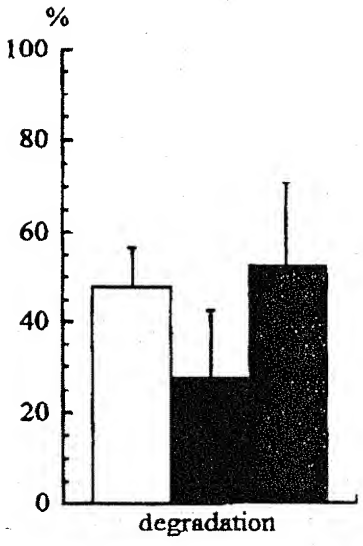


Fig. 12D



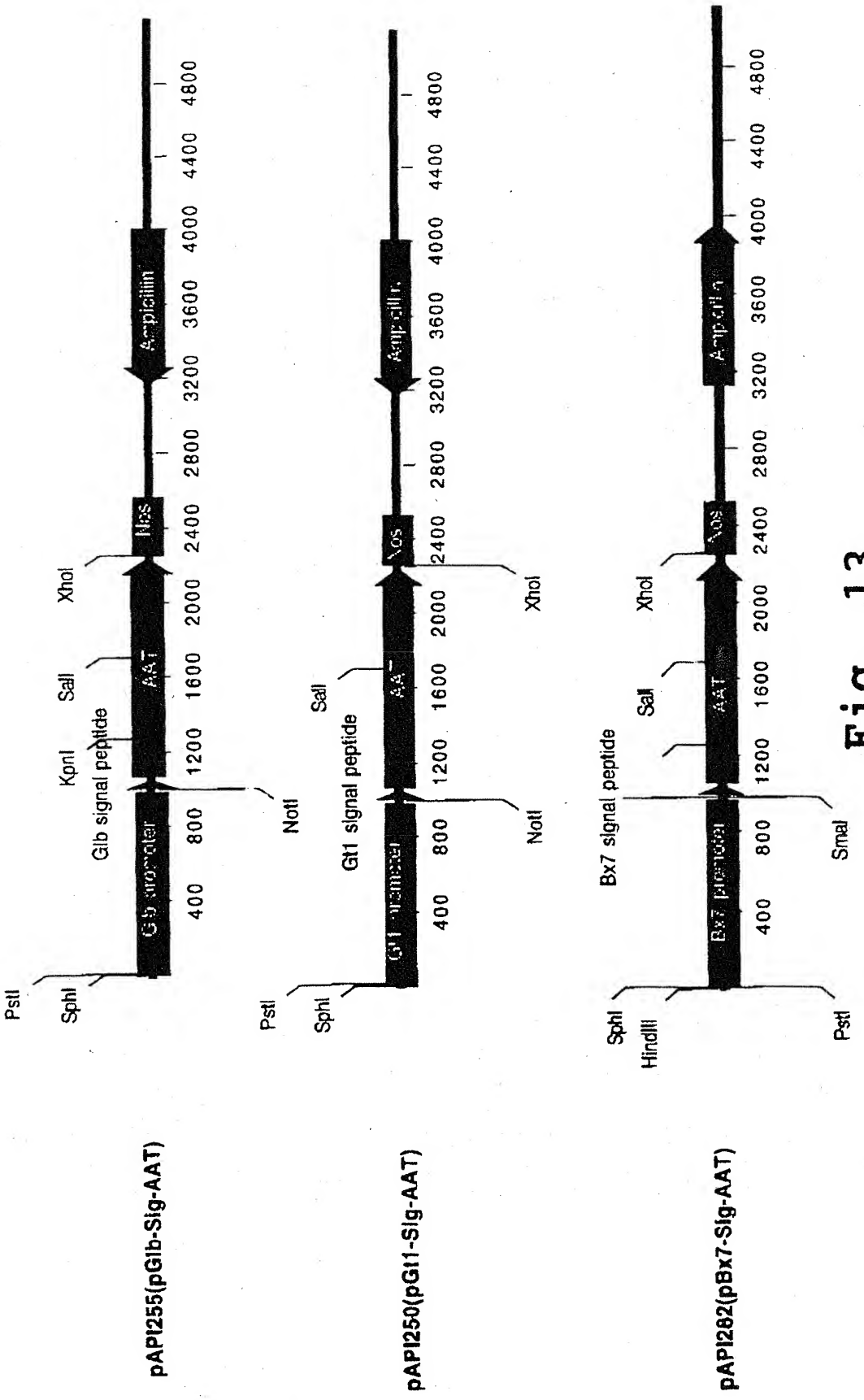


Fig. 13

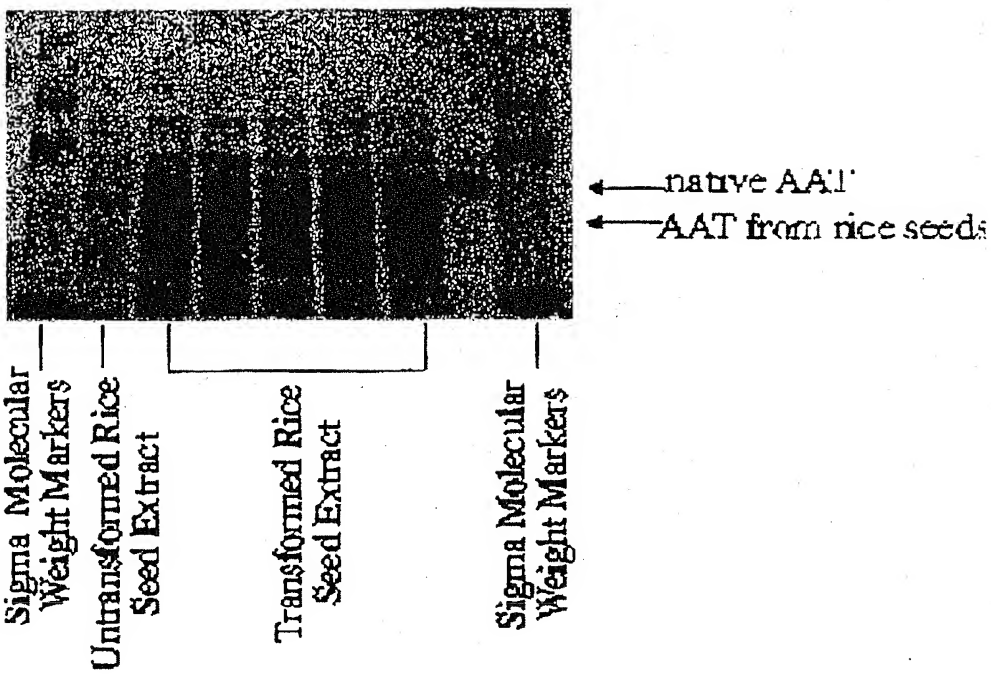

native AAT
AAT from rice seeds
Sigma Molecular Weight Markers
Untransformed Rice Seed Extract
Transformed Rice Seed Extract
Sigma Molecular Weight Markers

Fig. 14

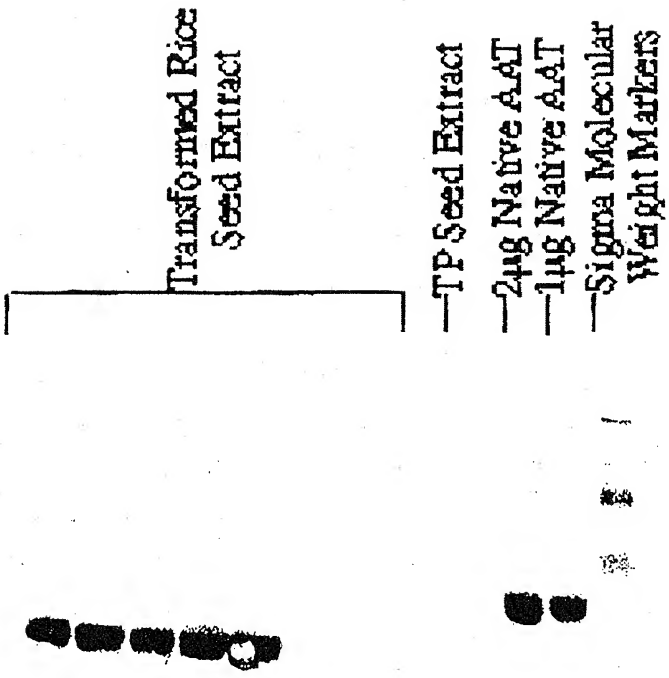


Fig. 15

Fig. 16A

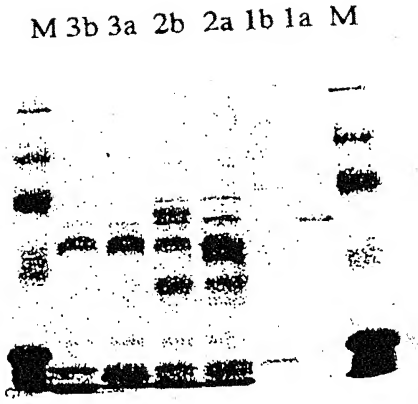
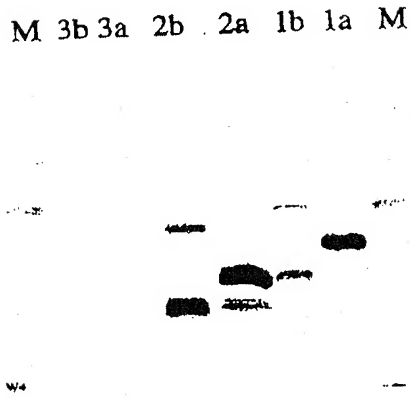


Fig. 16B



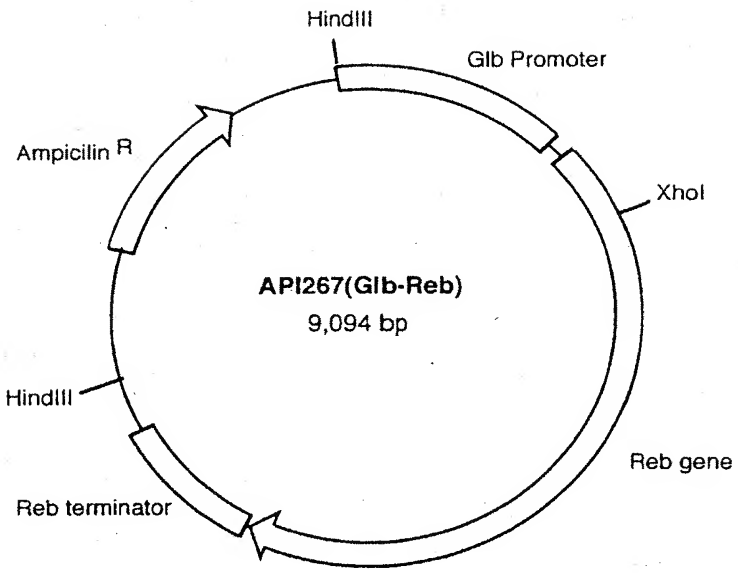

HindIII
Glb Promoter
XhoI
Reb gene
Reb terminator
HindIII
Ampicilin R
API267(Glb-Reb)
9,094 bp

Fig. 17A

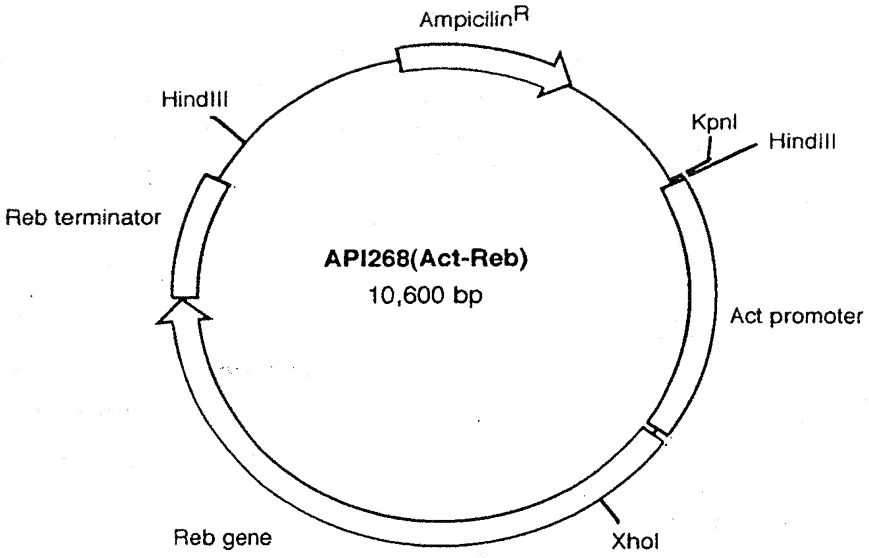


Fig. 17B

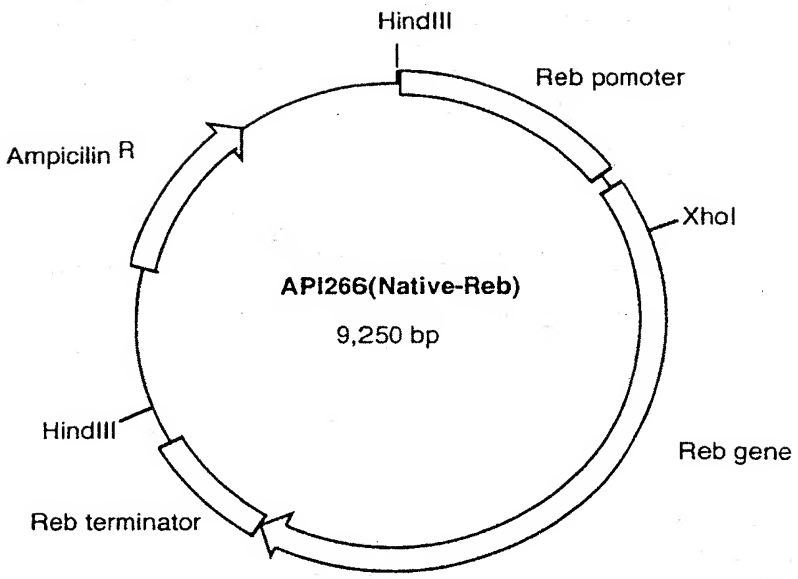

HindIII
Reb pomoter
Ampicilin R
XhoI
API266(Native-Reb)
9,250 bp
HindIII
Reb gene
Reb terminator

Fig. 17C

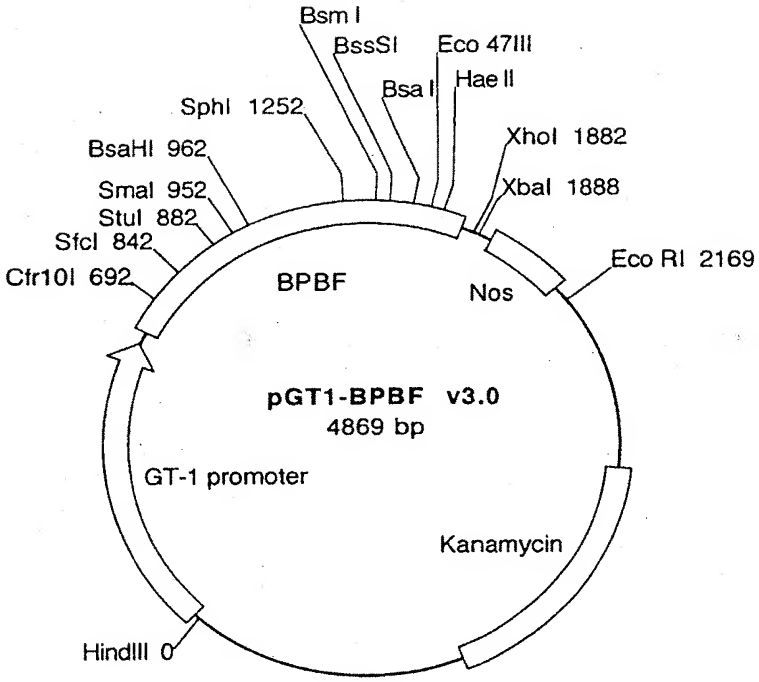

Bsm I
BssSI
Eco 47III
Hae II
Bsa I
SphI 1252
XhoI 1882
BsaHI 962
XbaI 1888
SmaI 952
StuI 882
SfcI 842
Cfr10I 692
Eco RI 2169
BPBF
Nos
pGT1-BPBF v3.0
4869 bp
GT-1 promoter
Kanamycin
HindIII 0

Fig. 18A

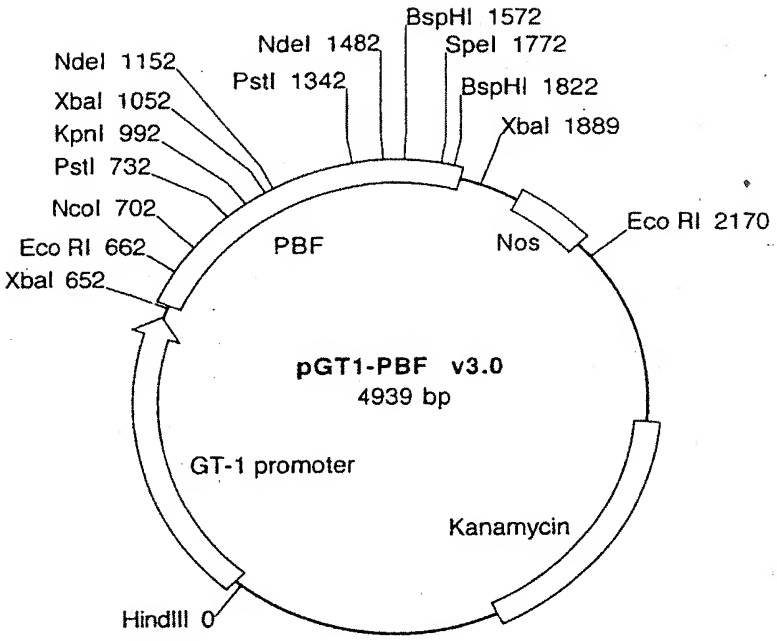

BspHI 1572
NdeI 1482
SpeI 1772
NdeI 1152
PstI 1342
BspHI 1822
XbaI 1052
XbaI 1889
KpnI 992
PstI 732
NcoI 702
Eco RI 2170
PBF
Nos
Eco RI 662
XbaI 652
pGT1-PBF v3.0
4939 bp
GT-1 promoter
Kanamycin
HindIII 0

Fig. 18B

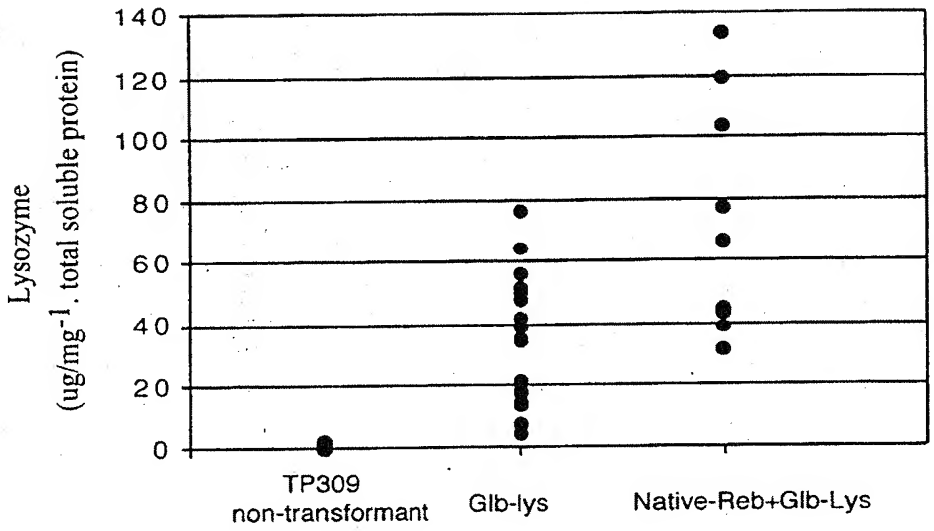

Lysozyme
(ug/mg-1. total soluble protein)
140
120
100
80
60
40
20
0
TP309
non-transformant
Glb-lys
Native-Reb+Glb-Lys

Fig. 19

Sequence Range Egfactor: 4 to 165

```
                  10          20          30          40          50
Egfactor     AAC TCC GAC TCG GAG TGC CCC CTC TCC CAC GAC GGT TAC TGC CTC CAC GAC GGG
              N   S   D   S   E   C   P   L   S   H   D   G   Y   C   L   H   D   G>

                        3360        3370        3380        3390        3400
Native Gene  AAT GGT GAC TCT GAA TGT CCC CTG TCC CAC GAT GGG TAC TGC CTC CAT GAT GGT
             ||     ||| ||  ||  ||  ||| ||  ||| ||| ||  ||  ||| ||| ||| ||  ||  ||
Egfactor     AAC TCC GAC TCG GAG TGC CCC CTC TCC CAC GAC GGT TAC TGC CTC CAC GAC GGG


              60          70          80          90          100         110
Egfactor     GTC TGC ATG TAC ATC GAG GCC CTC GAC AAG TAC GCC TGC AAC TGC GTC GTG GGC
              V   C   M   Y   I   E   A   L   D   K   Y   A   C   N   C   V   V   G>

                  3410        3420        3430        3440        3450
Native Gene  GTG TGC ATG TAT ATT GAA GCA TTG GAC AAG TAT GCA TGC AAC TGT GTT GTT GGC
             ||  ||| ||| ||  ||  ||  ||   |  ||| ||| ||  ||  ||| ||| ||  ||  ||  |||
Egfactor     GTC TGC ATG TAC ATC GAG GCC CTC GAC AAG TAC GCC TGC AAC TGC GTC GTG GGC


                  120         130         140         150         160
Egfactor     TAC ATC GGC GAG CGG TGC CAG TAC CGC GAC CTC AAG TGG TGG GAG CTG CGC TGA
              Y   I   G   E   R   C   Q   Y   R   D   L   K   W   W   E   L   R   *>

                3460        3470        3480        3490        3500
Native Gene  TAC ATC GGG GAG CGA TGT CAG TAC CGA GAC CTG AAG TGG TGG GAA CTG CGC
             ||| ||| ||  ||| ||  ||  ||| ||| ||  ||| ||  ||| ||| ||| ||  ||| |||
Egfactor     TAC ATC GGC GAG CGG TGC CAG TAC CGC GAC CTC AAG TGG TGG GAG CTG CGC TGA
```

Epidermal Growth Factor

Number of codons in mature peptide:	53
Number of codons changed:	27 (51%)
Number of nucleotides changed:	30 (19%)

Fig. 20

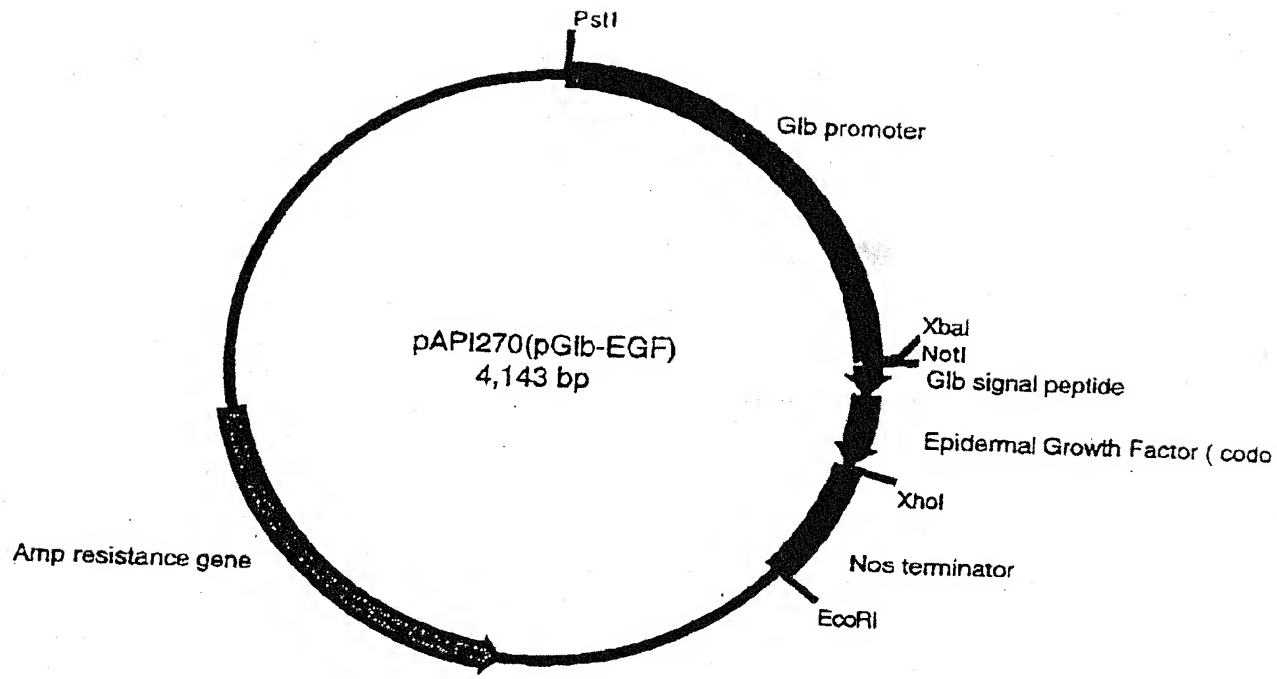

PstI
Glb promoter
pAPI270(pGlb-EGF)
4,143 bp
XbaI
NotI
Glb signal peptide
Epidermal Growth Factor (codo
XhoI
Nos terminator
EcoRI
Amp resistance gene

Fig. 21

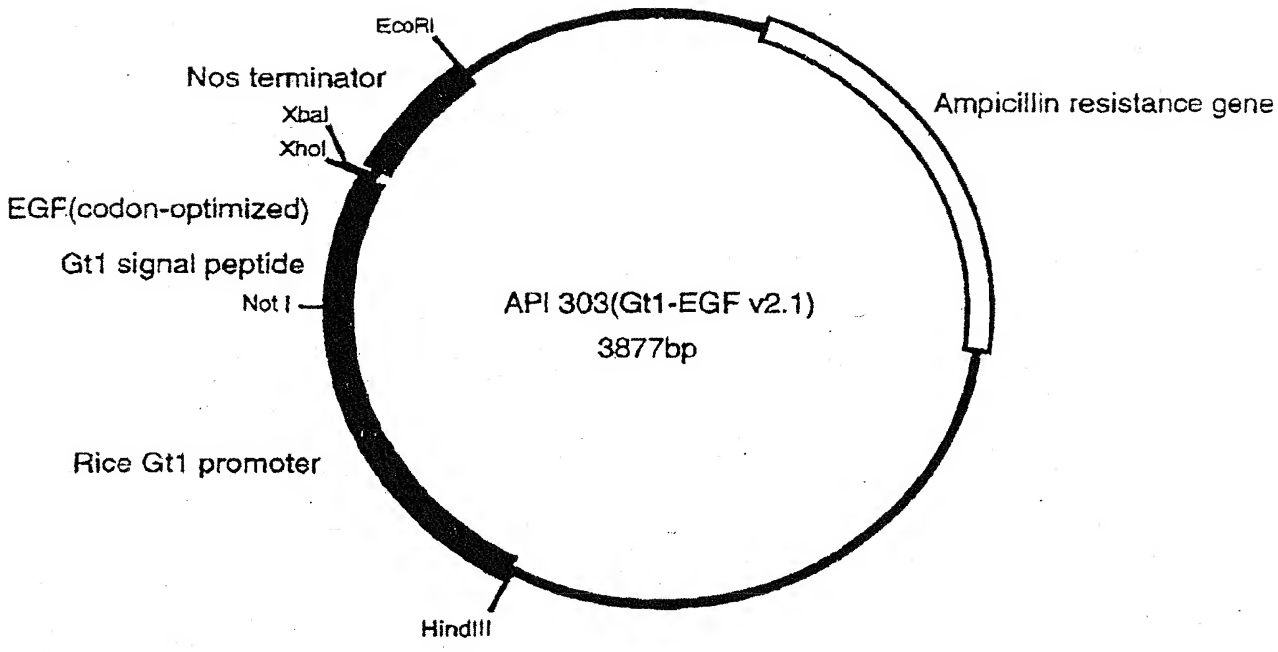

EcoRI
Nos terminator
XbaI
XhoI
EGF(codon-optimized)
Gt1 signal peptide
Not I
Rice Gt1 promoter
HindIII
Ampicillin resistance gene
API 303(Gt1-EGF v2.1)
3877bp

Fig. 22

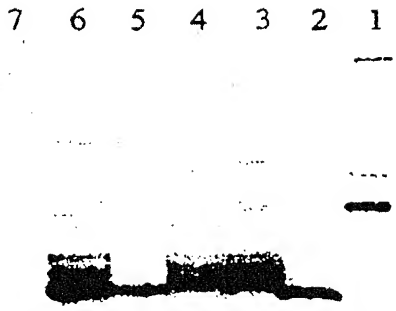


Fig. 23

Sequence Range: 4 to 216

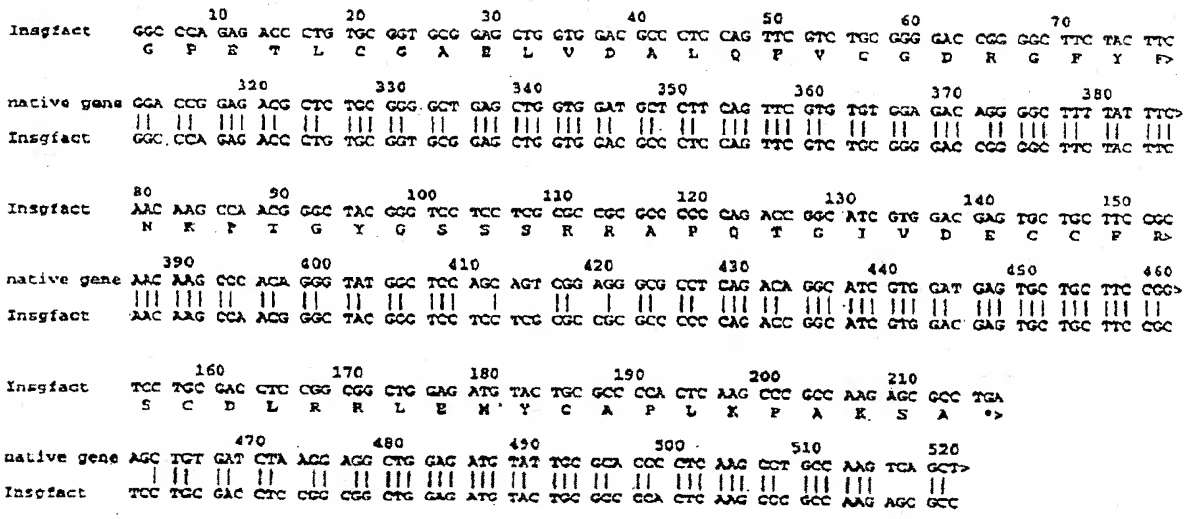


Fig. 24

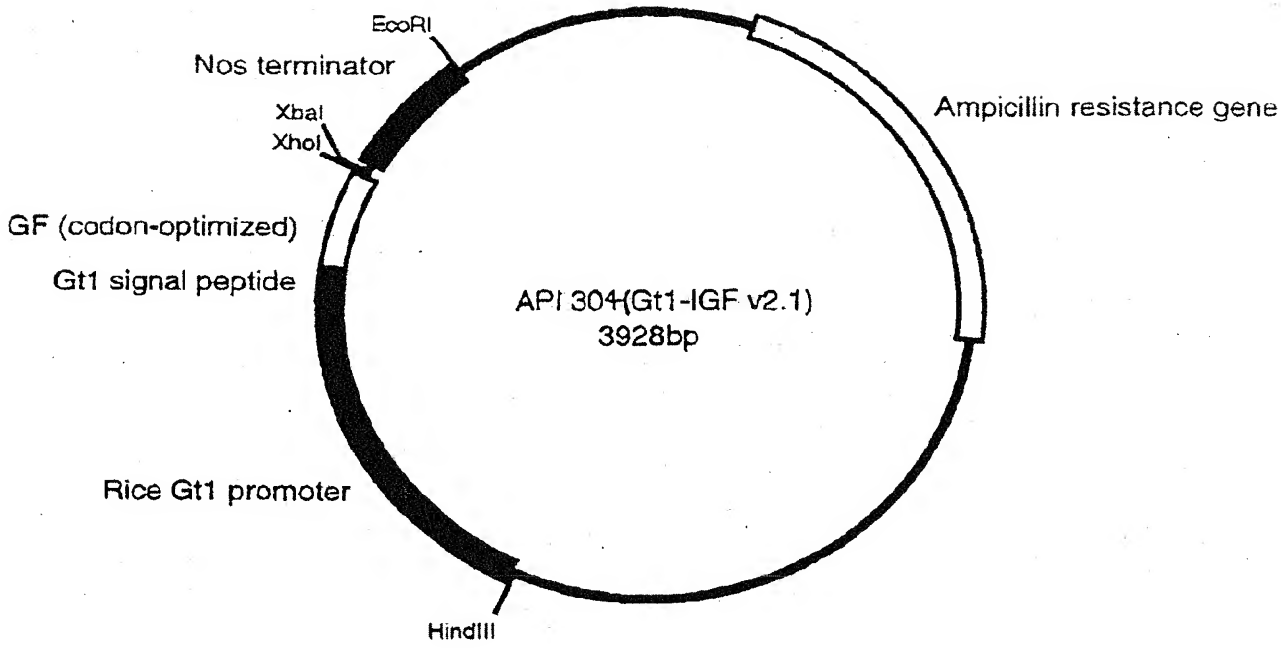


Fig. 25

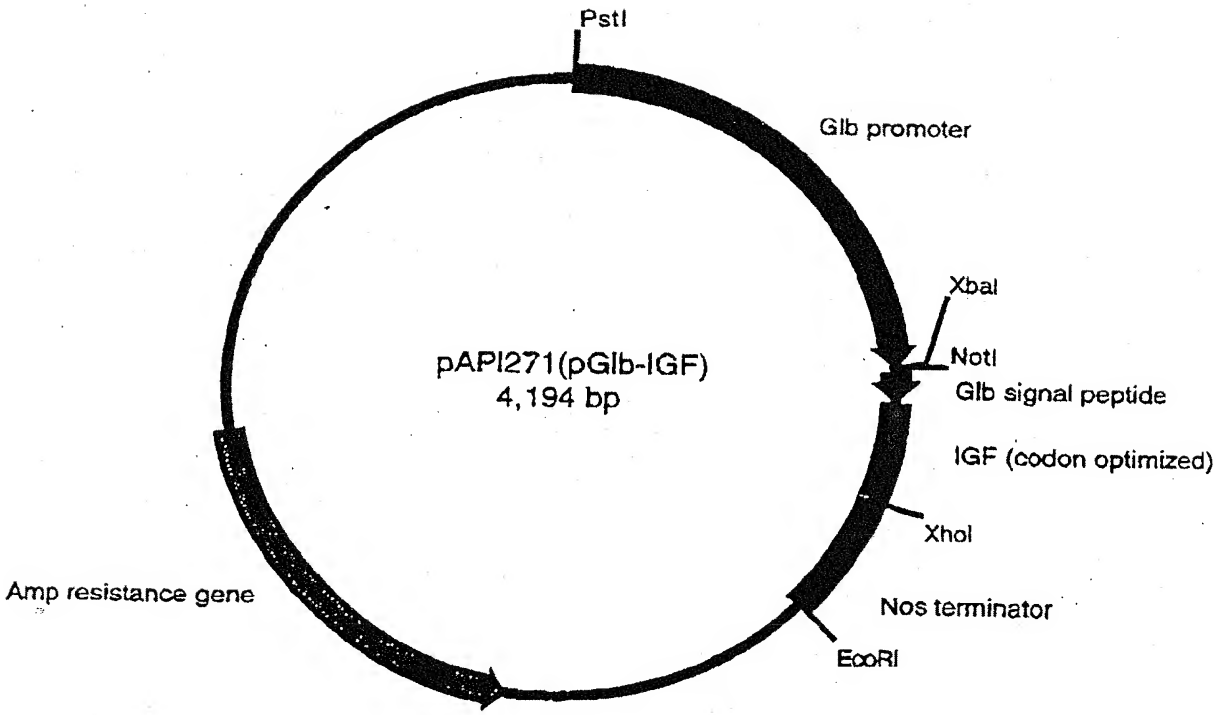

PstI
Glb promoter
XbaI
NotI
Glb signal peptide
IGF (codon optimized)
XhoI
Nos terminator
EcoRI
Amp resistance gene
pAPI271(pGlb-IGF)
4,194 bp

Fig. 26

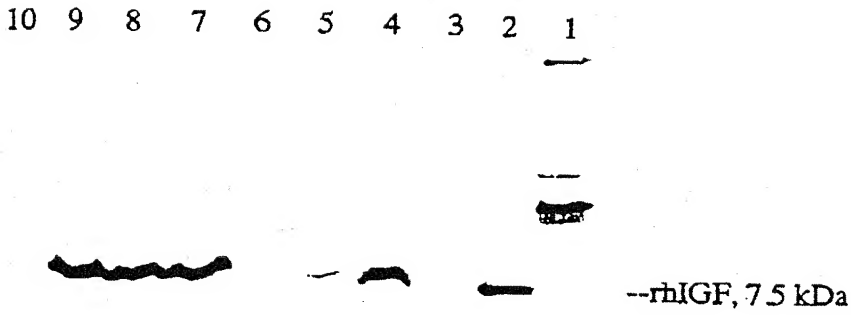


Fig. 27

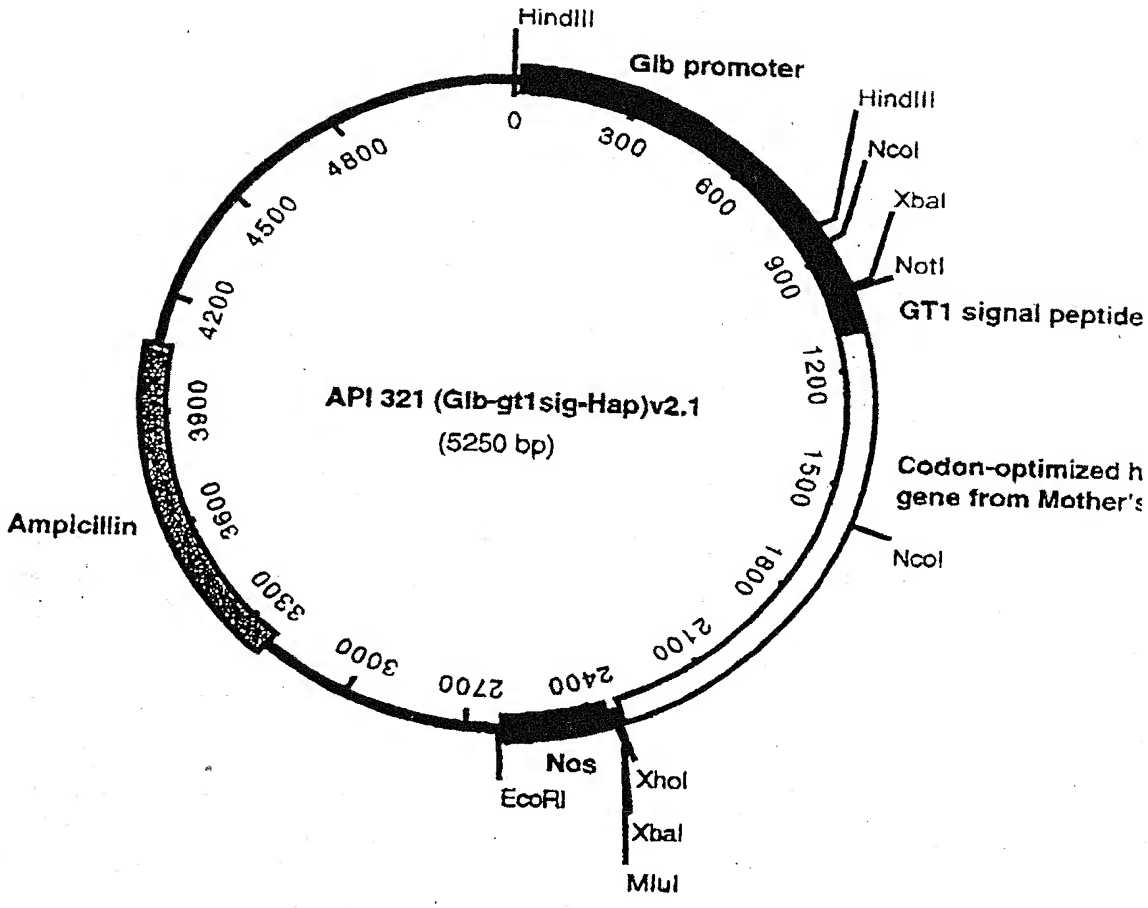

HindIII
Glb promoter
HindIII
NcoI
XbaI
NotI
GT1 signal peptide
Codon-optimized h
gene from Mother's
NcoI
XhoI
XbaI
MluI
Nos
EcoRI
Ampicillin
API 321 (Glb-gt1sig-Hap)v2.1
(5250 bp)
0
300
600
900
1200
1500
1800
2100
2400
2700
3000
3300
3600
3900
4200
4500
4800

Fig. 28

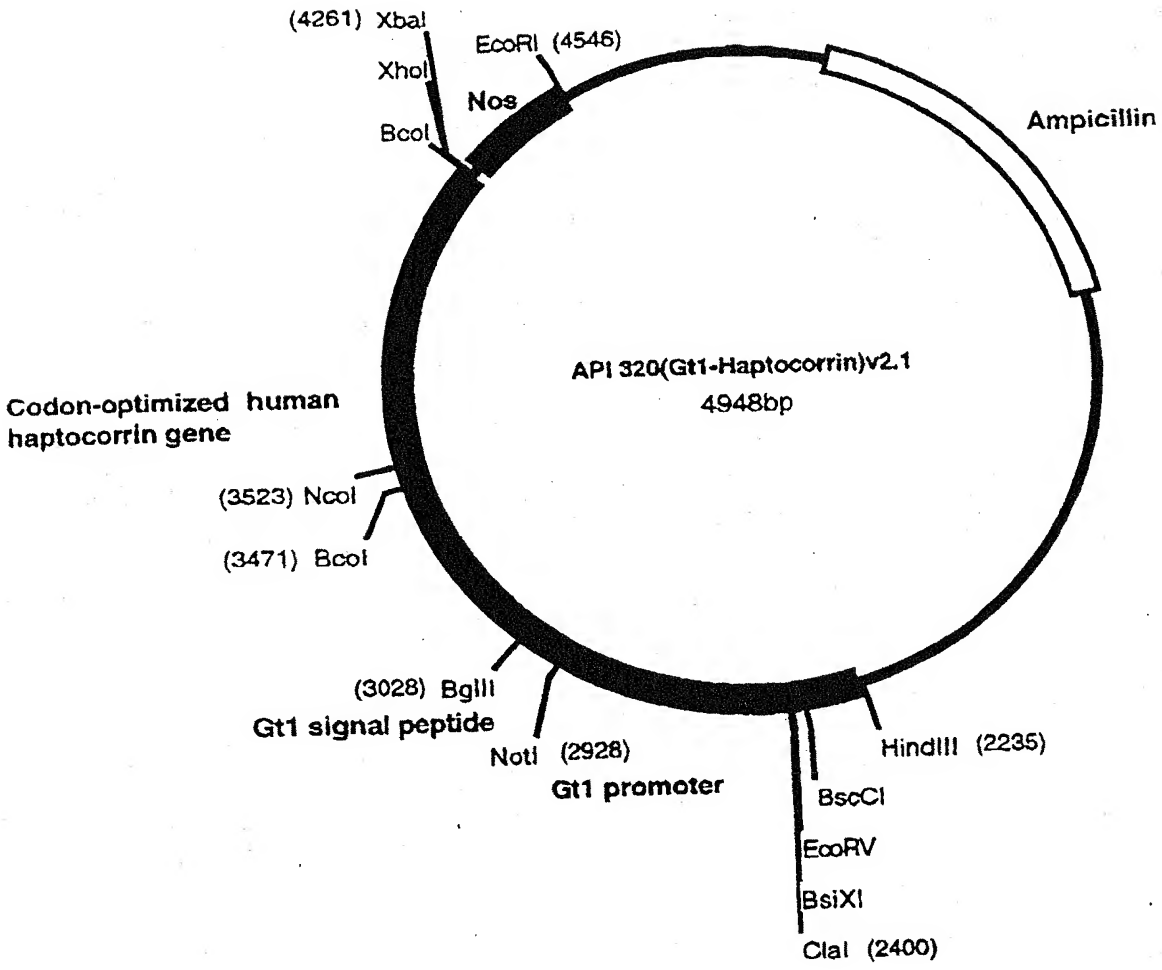


Fig. 29

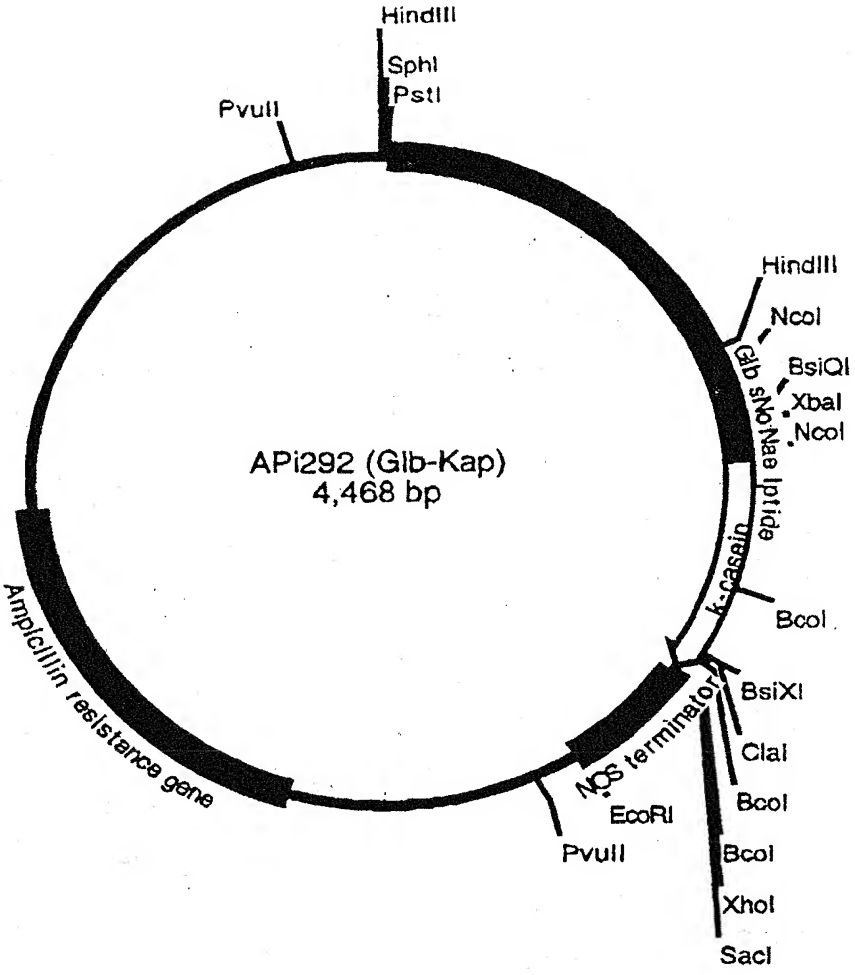

HindIII
SphI
PstI
PvuII
HindIII
NcoI
BsiQI
XbaI
NcoI
Glb sNo
Nae Iptide
k-casein
BcoI
BsiXI
ClaI
BcoI
BcoI
XhoI
SacI
NOS terminator
EcoRI
PvuII
Ampicillin resistance gene
APi292 (Glb-Kap)
4,468 bp

Fig. 30

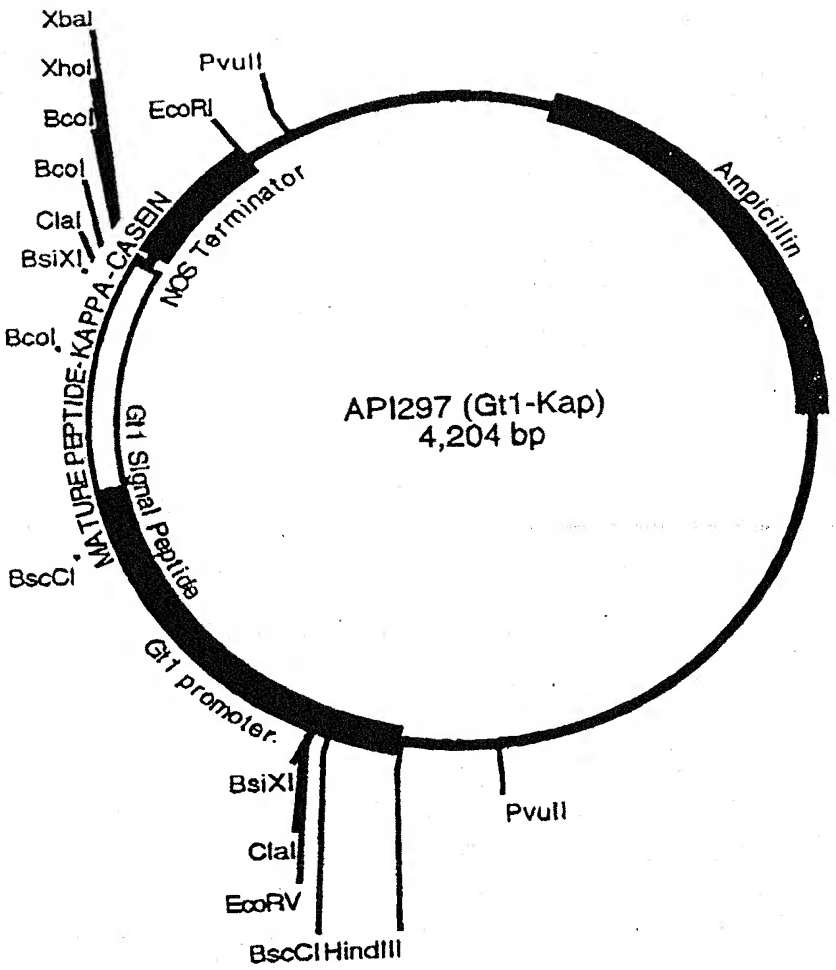

XbaI
XhoI
BcoI
BcoI
ClaI
BsiXI
BcoI
BsccI
PvuII
EcoRI
NOS Terminator
Ampicillin
MATURE PEPTIDE-KAPPA-CASEIN
Gt1 Signal Peptide
Gt1 promoter
AP1297 (Gt1-Kap)
4,204 bp
BsiXI
ClaI
EcoRV
BsccI HindIII
PvuII

Fig. 31

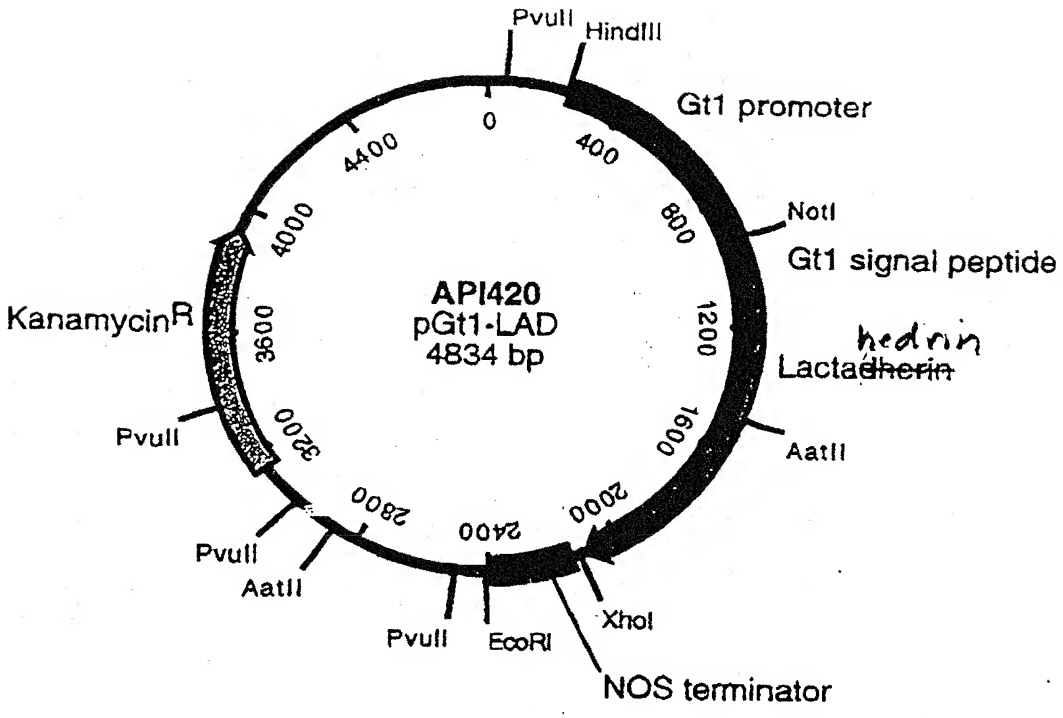

PvuII
HindIII
Gt1 promoter
NotI
Gt1 signal peptide
Lactadherin
AatII
XhoI
NOS terminator
EcoRI
PvuII
AatII
PvuII
PvuII
KanamycinR
API420
pGt1-LAD
4834 bp
0
400
800
1200
1600
2000
2400
2800
3200
3600
4000
4400

Fig. 32

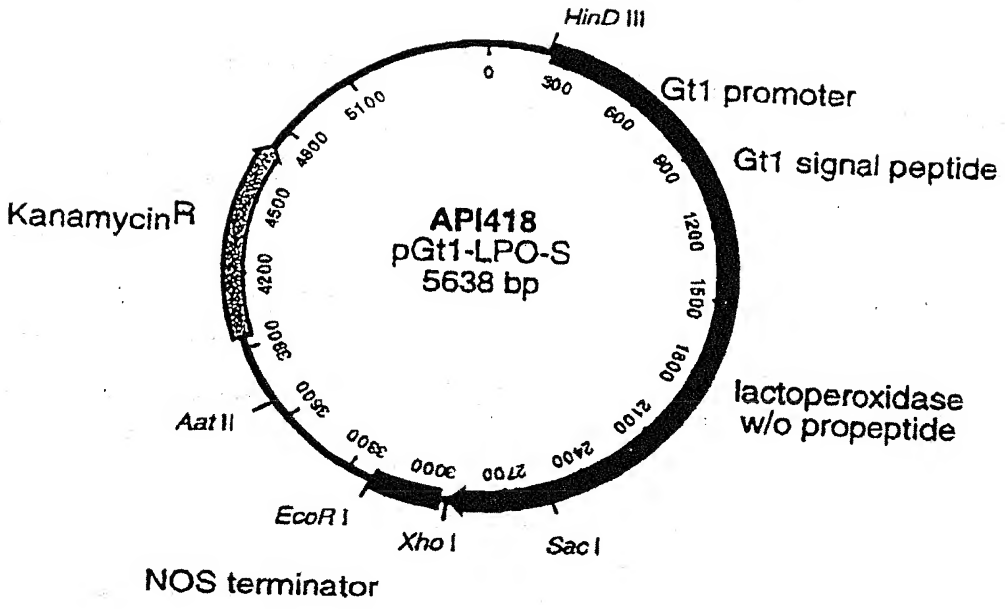

HinD III
Gt1 promoter
Gt1 signal peptide
lactoperoxidase
w/o propeptide
Sac I
Xho I
EcoR I
Aat II
NOS terminator
KanamycinR
API418
pGt1-LPO-S
5638 bp
0
300
600
900
1200
1500
1800
2100
2400
2700
3000
3300
3600
3900
4200
4500
4800
5100

Fig. 33

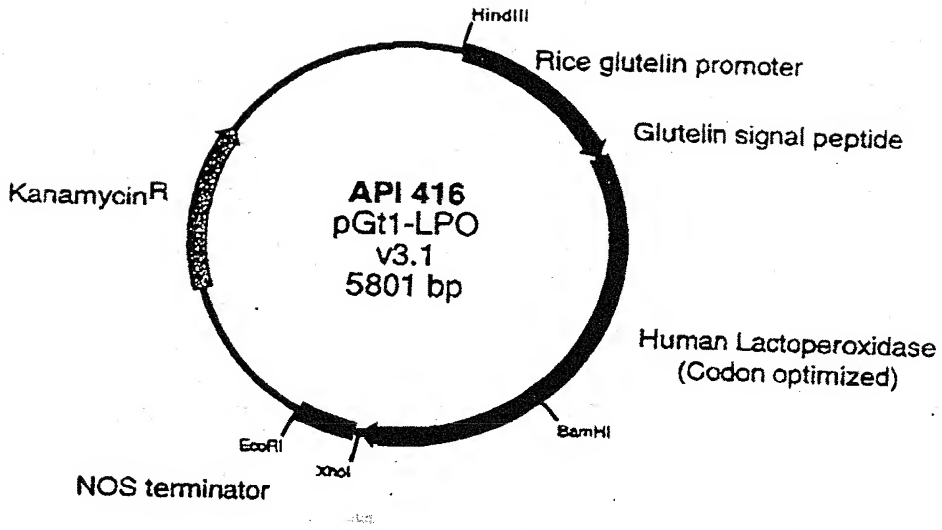

HindIII
Rice glutelin promoter
Glutelin signal peptide
KanamycinR
API 416
pGt1-LPO
v3.1
5801 bp
Human Lactoperoxidase
(Codon optimized)
BamHI
EcoRI
XhoI
NOS terminator

Fig. 34

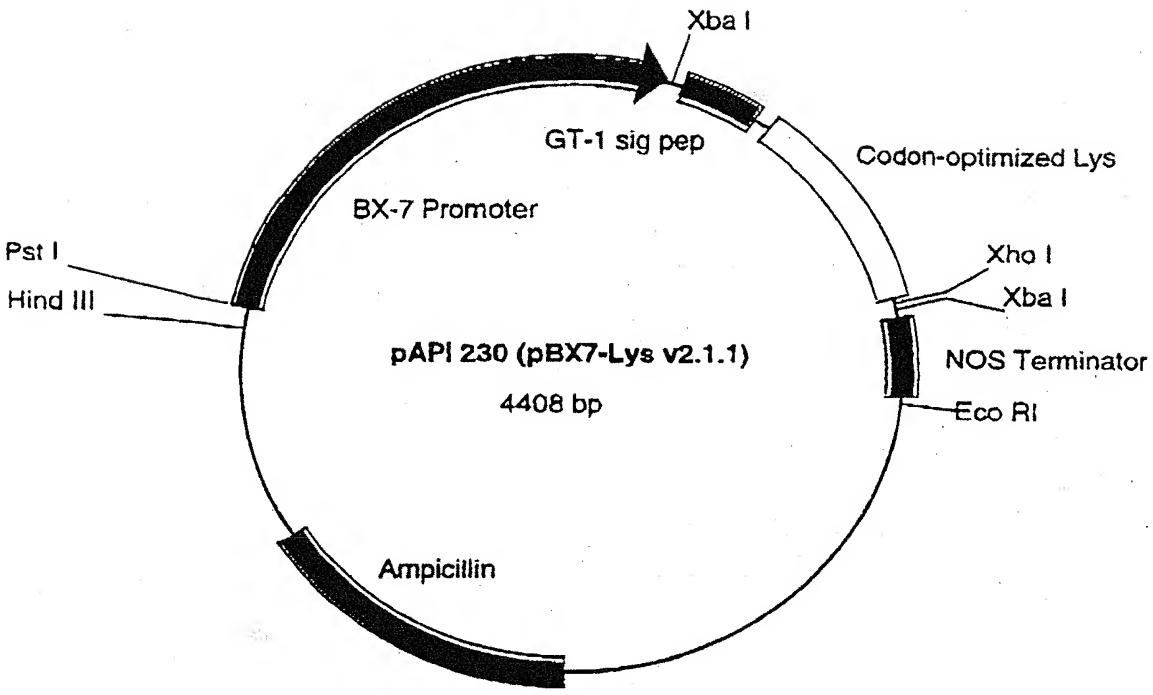

Xba I
GT-1 sig pep
Codon-optimized Lys
BX-7 Promoter
Pst I
Hind III
Xho I
Xba I
pAPI 230 (pBX7-Lys v2.1.1)
4408 bp
NOS Terminator
Eco RI
Ampicillin

Fig. 35

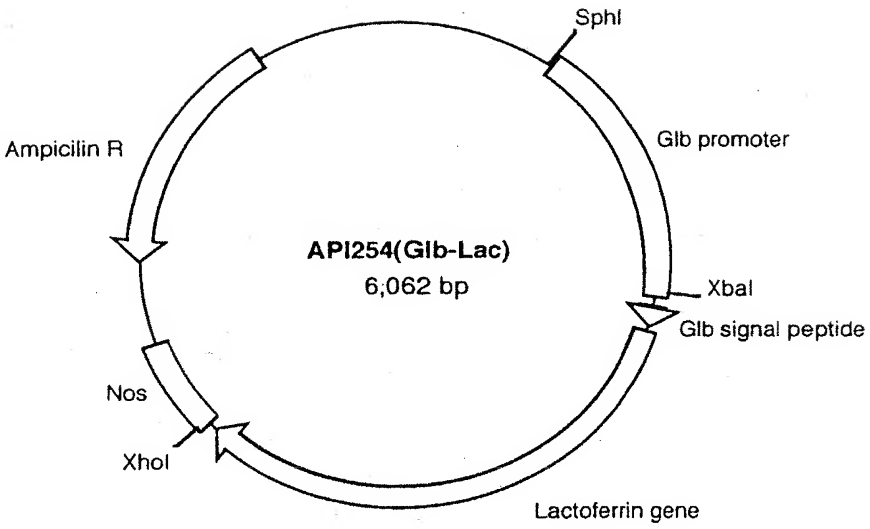

SphI
Glb promoter
Ampicilin R
API254(Glb-Lac)
6,062 bp
XbaI
Glb signal peptide
Nos
XhoI
Lactoferrin gene

Fig. 36A

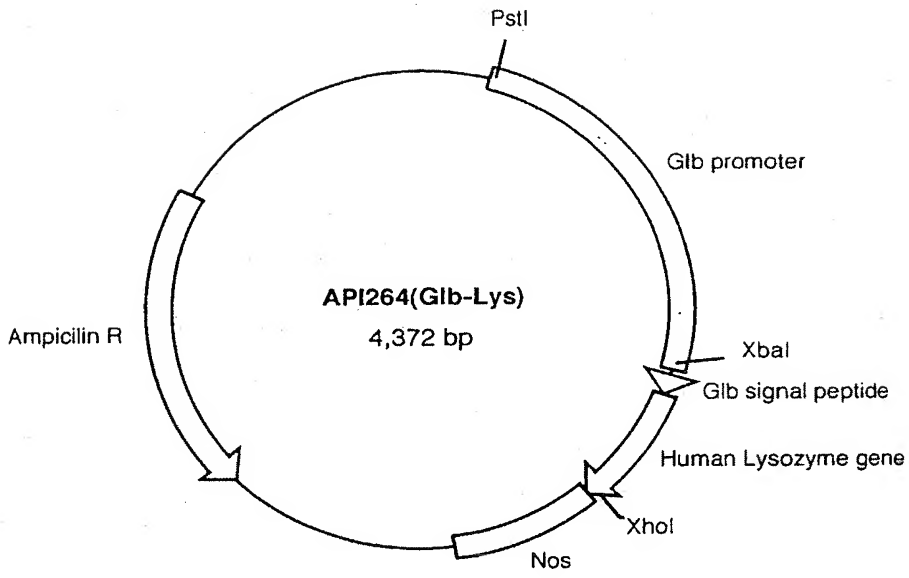

PstI
Glb promoter
API264(Glb-Lys)
4,372 bp
Ampicilin R
XbaI
Glb signal peptide
Human Lysozyme gene
XhoI
Nos

Fig. 36B

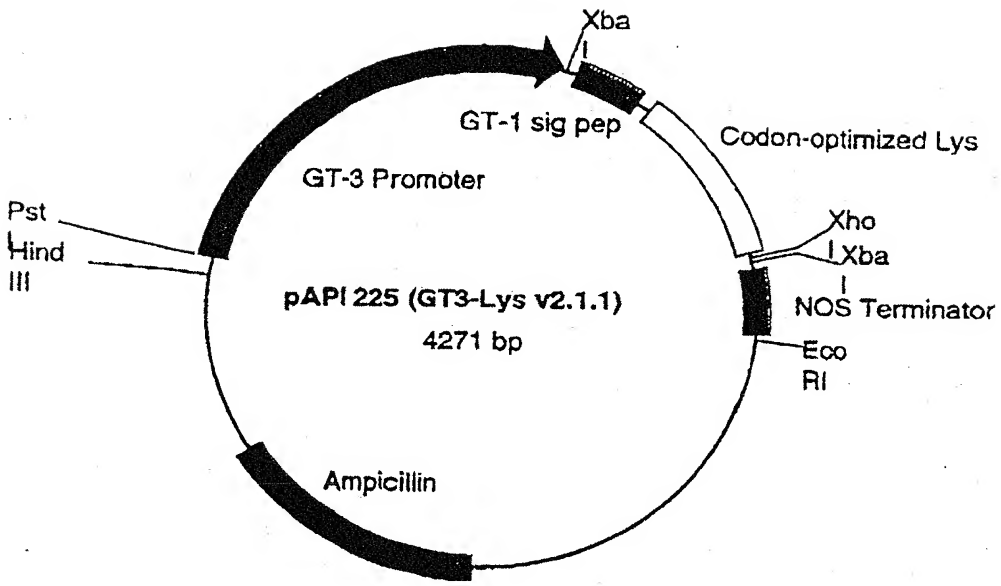

Xba
I
GT-1 sig pep
Codon-optimized Lys
GT-3 Promoter
Pst
I
Hind
III
Xho
I
Xba
I
NOS Terminator
pAPI 225 (GT3-Lys v2.1.1)
4271 bp
Eco
RI
Ampicillin

Fig. 37

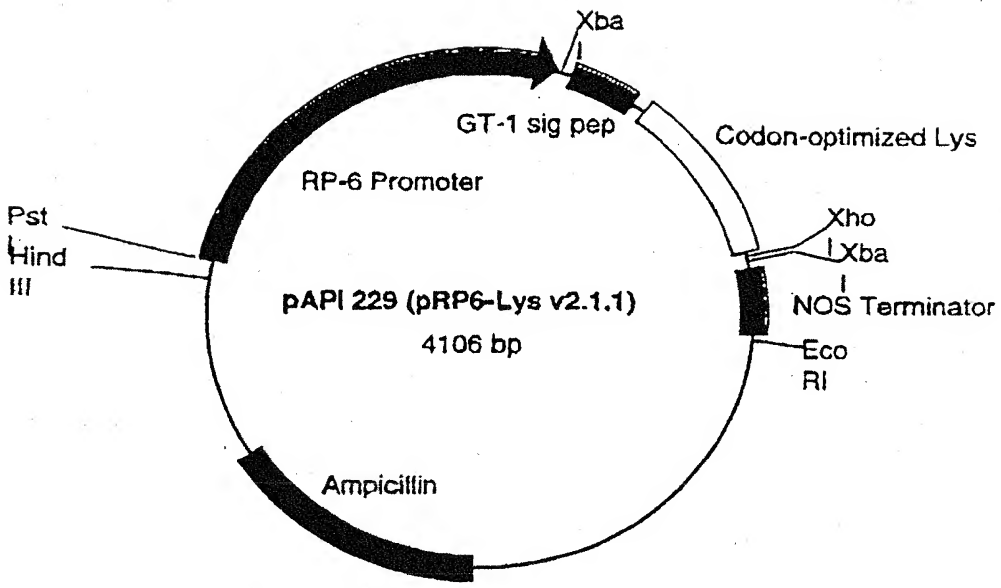

Xba
GT-1 sig pep
Codon-optimized Lys
RP-6 Promoter
Pst
Hind
III
Xho
Xba
pAPI 229 (pRP6-Lys v2.1.1)
4106 bp
NOS Terminator
Eco
RI
Ampicillin

Fig. 38

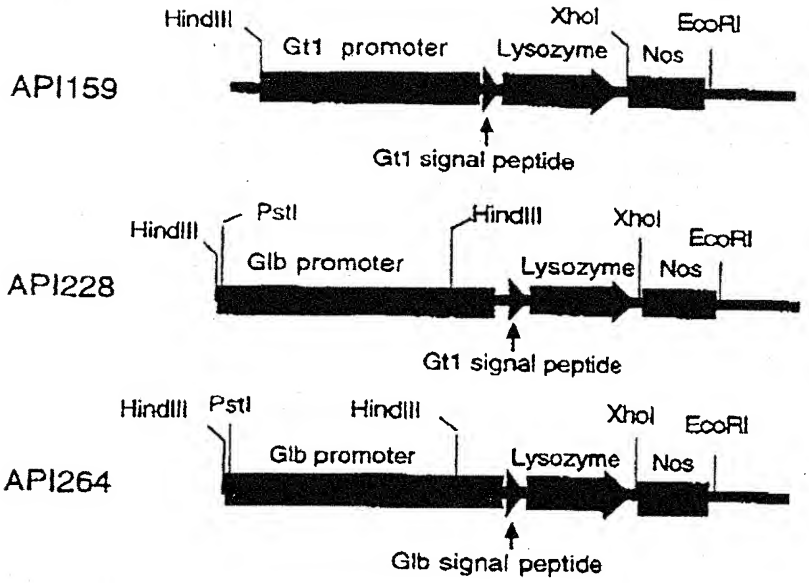


Fig. 39A

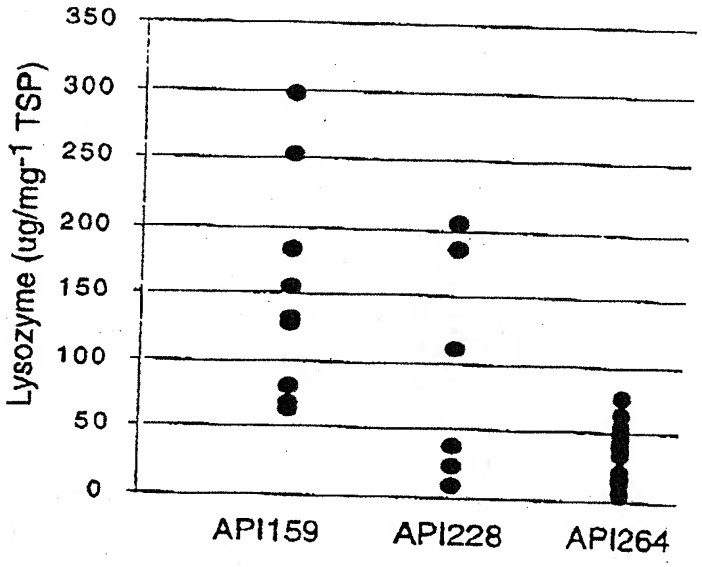


Fig. 39B

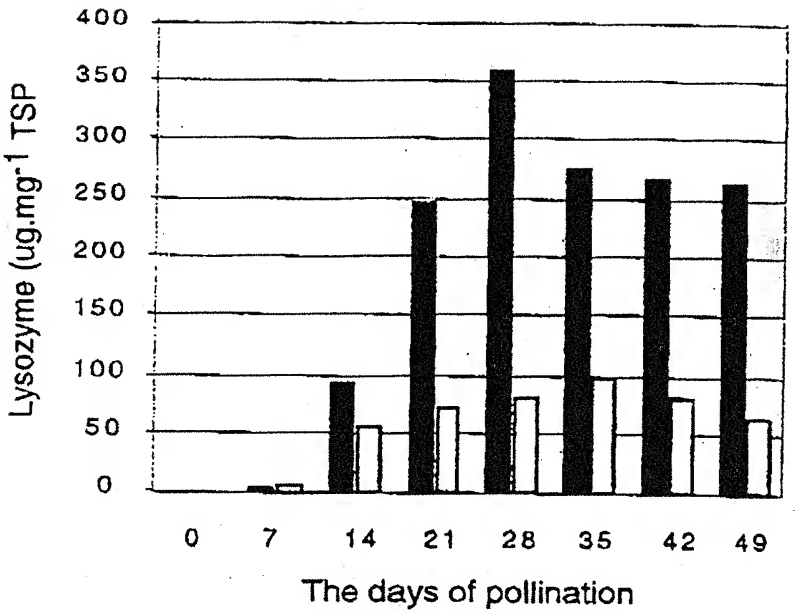


Fig. 40

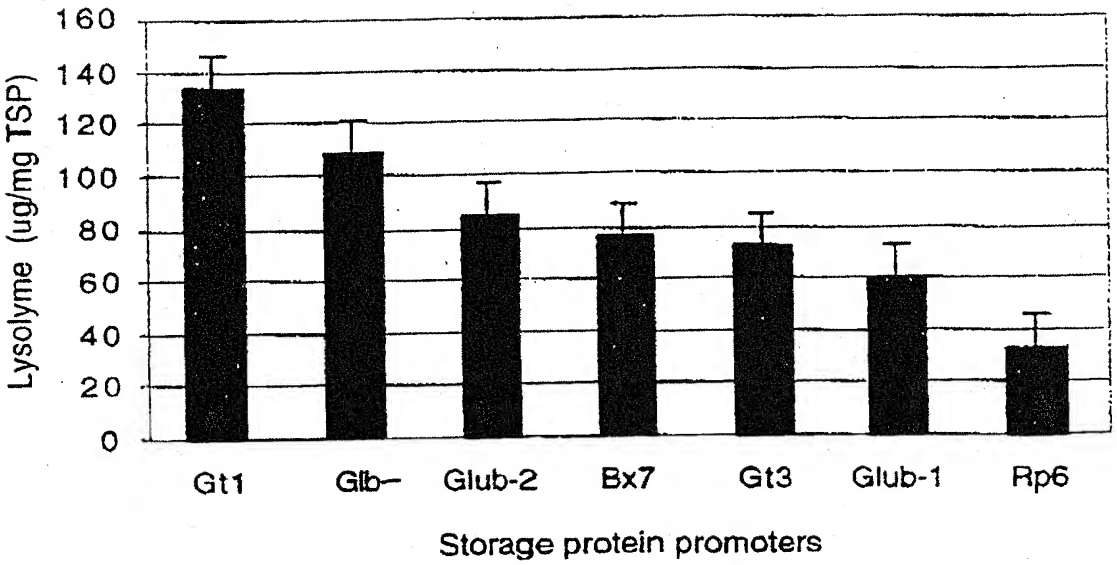

160
140
120
100
80
60
40
20
0
Lysolyme (ug/mg TSP)
Gt1
Glb–
Glub-2
Bx7
Gt3
Glub-1
Rp6
Storage protein promoters

Fig. 41

```
        10         20         30         40         50         60         70
GGTACCCATC TAATACATTA ATAACAAGAG AGAGAATGGA TAATGCAATT ATTTATTTTT ATGGGAGGCT
CCATGGGTAG ATTATGTAAT TATTGTTCTC TCTCTTACCT ATTACGTTAA TAAATAAAAA TACCCTCCGA

        80         90        100        110        120        130        140
ATATTTTTAT CGGATTTTAG TAAATAACGG GGCAATTCGG TACTTAGGTA AAGCTACGTA TGACTATCGC
TATAAAAATA GCCTAAAATC ATTTATTGCC CCGTTAAGCC ATGAATCCAT TTCGATGCAT ACTGATAGCG

       150        160        170        180        190        200        210
TACCGCTACG GTAGTTGAAT TGGAATTCTT CGATAGCATC TGTTGTGTTG TTGCAGTTAG GGTACTTGAA
ATGGCGATGC CATCAACTTA ACCTTAAGAA GCTATCGTAG ACAACACAAC AACGTCAATC CCATGAACTT

       220        230        240        250        260        270        280
TAGCTCCAGC CGTGAAAACG AGGGGTTTTC GCAGGTTTTA TAGGATTGCC AAGTTAGACT AGGGCAATTC
ATCGAGGTCG GCACTTTTGC TCCCCAAAAG CGTCCAAAAT ATCCTAACGG TTCAATCTGA TCCCGTTAAG

       290        300        310        320        330        340        350
ATGTTCACGG TATTGTGTAG TATATGAAAA AGGAGATCTC CCAAACAATT TATAATTTTG TATAAGGGAG
TACAAGTGCC ATAACACATC ATATACTTTT TCCTCTAGAG GGTTTGTTAA ATATTAAAAC ATATTCCCTC

                                                >AT-rich_region_
                                                |
       360        370        380        390        400        410        420
AAATCGAACT TGAGGTGTCT AATTCACCAA CCGAGCTACT CCCTCCGTTT CATATATGTA TATACATATA
TTTAGCTTGA ACTCCACAGA TTAAGTGGTT GGCTCGATGA GGGAGGCAAA GTATATACAT ATATGTATAT

       430        440        450        460        470        480        490
TACGTATATA TACGTATATA CACATATACG TATATACATA TATGGTATAT ACATATATAT ATATATATAT
ATGCATATAT ATGCATATAT GTGTATATGC ATATATGTAT ATACCATATA TGTATATATA TATATATATA

       500        510        520        530        540        550        560
ATATATATAT ATGTGTGTGT GTGTATGTGG GGTGGCAATG CTAAAAAGTT TTATAATATG AACGGATGAA
TATATATATA TACACACACA CACATACACC CCACCGTTAC GATTTTTCAA AATATTATAC TTGCCTACTT

       570        580        590        600        610        620        630
GTACTATCCA CTAAGTCCCT ATAGTTTTCT GGCACTGTGT AGTATACGAA TGCACAATTA TATCCATAAA
CATGATAGGT GATTCAGGGA TATCAAAAGA CCGTGACACA TCATATGCTT ACGTGTTAAT ATAGGTATTT

       640        650        660        670        680        690        700
ATTGATATTA TATATTCGTC GCGACGAAAA TAAAGACATA ATATTCGGTA TACCATTTAT CCACGATATA
TAACTATAAT ATATAAGCAG CGCTGCTTTT ATTTCTGTAT TATAAGCCAT ATGGTAAATA GGTGCTATAT

       710        720        730        740        750        760        770
TCTAAATTCC ACTGATATAT CTAAATTCCA CTTGATCCCT TTTATGGATA AATTCTGGAT AACAATTACT
AGATTTAAGG TGACTATATA GATTTAAGGT GAACTAGGGA AAATACCTAT TTAAGACCTA TTGTTAATGA

       780        790        800        810        820        830        840
ACCAGCAGTA TATCCTACTA TCAGCGCACT GCACACCAAA CTACCCTCAC CCAGTAGTTA CAAACGCATA
TGGTCGTCAT ATAGGATGAT AGTCGCGTGA CGTGTGGTTT GATGGGAGTG GGTCATCAAT GTTTGCGTAT
```

Fig. 42A

```
       850        860        870        880        890        900        910
TTTTGCCGTT AGTTAATTAT TATCCGGTAA AGAAGGTAAA GAAGATTGGT AGTAATCCAA AATTTTCCCA
AAAACGGCAA TCAATTAATA ATAGGCCATT TCTTCCATTT CTTCTAACCA TCATTAGGTT TTAAAAGGGT

       920        930        940        950        960        970        980
ACCCCAACCT CGGAACAAAA ACCGCGTAGT ATTTGTCGTA ACCAGGAGCA TCCGAGTCAT TAATTTACAC
TGGGGTTGGA GCCTTGTTTT TGGCGCATCA TAAACAGCAT TGGTCCTCGT AGGCTCAGTA ATTAAATGTG

                >Transcription_start_site
>CAAG_site
                                  |                                       |
       990       1000       1010       1020       1030       1040       1050
CCAAACACAA AAAATTAGCA GCACGCAGCC GCCTTCCCAA TCCTCTCCTC TCTCCTCTCC TCTTCTCCAA
GGTTTGTGTT TTTTAATCGT CGTGCGTCGG CGGAAGGGTT AGGAGAGGAG AGAGGAGAGG AGAAGAGGTT

      1060       1070       1080       1090       1100       1110       1120
GCGGCAATTC GCGCGAGGTT TTCTCCGATC AAACCCTCGA ATCCCCCCCT CGCGAATCCA TCGGAGGGTA
CGCCGTTAAG CGCGCTCCAA AAGAGGCTAG TTTGGGAGCT TAGGGGGGGA GCGCTTAGGT AGCCTCCCAT

      1130       1140       1150       1160          1170          1180
GCCCCGCGAT CCGCGTCGGC GAGAGCGGAT TCCGATTCCG CG ATG GAG CGG GTG TTC TCC GTG
CGGGGCGCTA GGCGCAGCCG CTCTCGCCTA AGGCTAAGGC GC TAC CTC GCC CAC AAG AGG CAC
                                                M   E   R   V   F   S   V>
                                               ___a___a__EXON1____a___a___>

     1190          1200          1210          1220          1230          1240
GAG GAG ATC TCC GAC CCA TTC TGG GTC CCG CCT CCG CCG CCG CAG TCG GCG GCG GCG
CTC CTC TAG AGG CTG GGT AAG ACC CAG GGC GGA GGC GGC GGC GTC AGC CGC CGC CGC
 E   E   I   S   D   P   F   W   V   P   P   P   P   P   Q   S   A   A   A>
___a___a___a___a___a___a___a___a__EXON1____a___a___a___a___a___a___a___a___>

         1250          1260          1270          1280          1290
GCC CAG CAG CAG GGC GGC GGC GGC GTG GCT TCG GGA GGT GGT GGT GGT GTA GCG GGG
CGG GTC GTC GTC CCG CCG CCG CCG CAC CGA AGC CCT CCA CCA CCA CCA CAT CGC CCC
 A   Q   Q   Q   G   G   G   G   V   A   S   G   G   G   G   G   V   A   G>
___a___a___a___a___a___a___a___a__EXON1____a___a___a___a___a___a___a___a___>

1300          1310          1320          1330          1340          1350
GGC GGC GGC GGC GGG AAC GCG ATG AAC CGG TGC CCG TCG GAG TGG TAC TTC CAG AAG
CCG CCG CCG CCG CCC TTG CGC TAC TTG GCC ACG GGC AGC CTC ACC ATG AAG GTC TTC
 G   G   G   G   G   N   A   M   N   R   C   P   S   E   W   Y   F   Q   K>
___a___a___a___a___a___a___a___a__EXON1____a___a___a___a___a___a___a___a___>

   1360          1370          1380          1390          1400          1410
TTT CTG GAG GAG GCG GTG CTC GAT AGC CCC GTC CCG AAC CCT AGC CCG AGG GCC GAA
AAA GAC CTC CTC CGC CAC GAG CTA TCG GGG CAG GGC TTG GGA TCG GGC TCC CGG CTT
 F   L   E   E   A   V   L   D   S   P   V   P   N   P   S   P   R   A   E>
___a___a___a___a___a___a___a___a__EXON1____a___a___a___a___a___a___a___a___>
```

Fig. 42B

```
       1420          1430          1440          1450          1460
GCG GGA GGG ATC AGG GGC GCA GGA GGG GTG GTG CCG GTC GAT GTT AAG CAG CCG CAG
CGC CCT CCC TAG TCC CCG CGT CCT CCC CAC CAC GGC CAG CTA CAA TTC GTC GGC GTC
 A   G   G   I   R   G   A   G   G   V   V   P   V   D   V   K   Q   P   Q>
___a___a___a___a___a___a___a___a__EXON1____a___a___a___a___a___a___a___a___>

       1470          1480          1490          1500          1510
1520
CTC TCG GCG GCG GCG ACG ACG AGC GCG GTG GTG GAC CCC GTG GAG TAC AAC GCG ATG
GAG AGC CGC CGC CGC TGC TGC TCG CGC CAC CAC CTG GGG CAC CTC ATG TTG CGC TAC
 L   S   A   A   A   T   T   S   A   V   V   D   P   V   E   Y   N   A   M>
___a___a___a___a___a___a___a___a___a__EXON1____a___a___a___a___a___a___a___a___>

  1530          1540          1550          1560          1570          1580
CTG AAG CAG AAG CTG GAG AAG GAC CTC GCC GCG GTC GCC ATG TGG AGG GTACAGC
GAC TTC GTC TTC GAC CTC TTC CTG GAG CGG CGC CAG CGG TAC ACC TCC CATGTCG
 L   K   Q   K   L   E   K   D   L   A   A   V   A   M   W   R>
___a___a___a___a___a___a___a____EXON1__a___a___a___a___a___a___a___a___>

       1590          1600          1610          1620          1630          1640          1650
CATTCTCCCC CCCTCTAGTA CTCGAGAGCT TACTGAGATC GGCAATGCTA GCTACTGTTT GCATCGAATG
GTAAGAGGGG GGGAGATCAT GAGCTCTCGA ATGACTCTAG CCGTTACGAT CGATGACAAA CGTAGCTTAC

       1660          1670          1680          1690          1700          1710          1720
TTTATAGGTA TTTAGATCGG GCATTTCTAT AGACCAATGG CGTCCATGGT CTTGCAATGC GCTCTGTTGA
AAATATCCAT AAATCTAGCC CGTAAAGATA TCTGGTTACC GCAGGTACCA GAACGTTACG CGAGACAACT

       1730          1740          1750          1760          1770          1780          1790
GTGTCGGTGG TTGGTTCGAC TCATAGTATG TAGGGTTGTG CGTATGTACA AACGGAAGCT TCATAGACCT
CACAGCCACC AACCAAGCTG AGTATCATAC ATCCCAACAC GCATACATGT TTGCCTTCGA AGTATCTGGA

       1800          1810          1820          1830          1840          1850          1860
CGGTATTGAG ATTGCGATAT CGATGCAACC TGCGAATTGG CGATGTAATC AGTCATATTC TTACTAAACT
GCCATAACTC TAACGCTATA GCTACGTTGG ACGCTTAACC GCTACATTAG TCAGTATAAG AATGATTTGA

       1870          1880          1890          1900          1910          1920          1930
GCGAGACAGT GGTTTGTTTG CAATTGCAAT ATTTTTGTAT GGGGCTGCTT AAACTGTCAT TGCCTTTTTA
CGCTCTGTCA CCAAACAAAC GTTAACGTTA TAAAAACATA CCCCGACGAA TTTGACAGTA ACGGAAAAAT


       1940          1950          1960          1970          1980          1990          2000
GATTGGCAAT ATGTGACTTT ATGCAAGTAT TTGATTGGGC GGATCCAGGA ACAAAAAGTT GGGGGGATTC
CTAACCGTTA TACACTGAAA TACGTTCATA AACTAACCCG CCTAGGTCCT TGTTTTTCAA CCCCCCTAAG

       2010          2020          2030          2040          2050          2060          2070
AACATACCGA GTACACTGGC ATAAACACAT CATCTCAGTA TTAAACTATG CTAAAATGCT ATTAAGAGAC
TTGTATGGCT CATGTGACCG TATTTGTGTA GTAGAGTCAT AATTTGATAC GATTTTACGA TAATTCTCTG
```

Fig. 42C

```
      2080        2090        2100        2110        2120        2130        2140
CTTTAGCACC TCTTATCTTA TCAACCATGG TGAAAAAATT GAAGGGGGGA CTCAGGGGGG TATCCATGGG
GAAATCGTGG AGAATAGAAT AGTTGGTACC ACTTTTTTAA CTTCCCCCCT GAGTCCCCCC ATAGGTACCC

      2150        2160        2170        2180        2190        2200        2210
TCCGATGGGT GCAGGGGGGA CTGAGTCCCC CCTGCACCCA CGTTGAATCC GCCCTGGCAT GCGTATAAGC
AGGCTACCCA CGTCCCCCCT GACTCAGGGG GGACGTGGGT GCAACTTAGG CGGGACCGTA CGCATATTCG

      2220        2230        2240        2250        2260        2270        2280
TGTCACAGCC ATTTCTAGGT GCTTGTGCTT AGTTGGGTGA TGTCAGCTTA ATTTGTCTTT TCTATGTCGT
ACAGTGTCGG TAAAGATCCA CGAACACGAA TCAACCCACT ACAGTCGAAT TAAACAGAAA AGATACAGCA

      2290        2300        2310        2320        2330        2340        2350
CATCGATTTT CTAAGAAACG AAAAATAGCC TATTTATGTG CTCCAGAATT TGATGATCCC TGGCCCTTCA
GTAGCTAAAA GATTCTTTGC TTTTTATCGG ATAAATACAC GAGGTCTTAA ACTACTAGGG ACCGGGAAGT

      2360        2370        2380        2390        2400        2410        2420
TTTGCTGAAA TTAGCCTATT TGTTGGTTGC CCTTCAGTTT TTTCCCAGCT TATGTTGTTG CAATGTGTGG
AAACGACTTT AATCGGATAA ACAACCAACG GGAAGTCAAA AAAGGGTCGA ATACAACAAC GTTACACACC

      2430        2440        2450        2460        2470        2480        2490
CTATGCCTCG TTTTGTGCCC TATAATTTAT TATTTGCAAT TCATTTTTGT ACATGACTTA AAATGACACT
GATACGGAGC AAAACACGGG ATATTAAATA ATAAACGTTA AGTAAAAACA TGTACTGAAT TTTACTGTGA

      2500        2510        2520        2530        2540        2550        2560
AGAGCAACAT GCACTGATTG GTTATCCTAT AATCATTTAT GTAGTTCTGT TCATTTTATC ATGCTAGCTC
TCTCGTTGTA CGTGACTAAC CAATAGGATA TTAGTAAATA CATCAAGACA AGTAAAATAG TACGATCGAG

      2570        2580        2590        2600        2610        2620
ATGTCATTTT CATCTTCAG GCC TCT GGC ACA GTT CCA CCT GAG CGT CCT GGA GCT GGT TCA
TACAGTAAAA GTAGAAGTC CGG AGA CCG TGT CAA GGT GGA CTC GCA GGA CCT CGA CCA AGT
                    A   S   G   T   V   P   P   E   R   P   G   A   G   S>
                   ___b___b___b___b___b____EXON2__b___b___b___b___b___b___>

      2630        2640        2650        2660        2670        2680
TCC TTG CTG AAT GCA GAT GTT TCA CAC ATA GGC GCT CCT AAT TCC ATC GGA GGTACTTA
AGG AAC GAC TTA CGT CTA CAA AGT GTG TAT CCG CGA GGA TTA AGG TAG CCT CCATGAAT
 S   L   L   N   A   D   V   S   H   I   G   A   P • N   S   I   G>
___b___b___b___b___b___b___b__EXON2____b___b___b___b___b___b___b___>

      2690        2700        2710        2720        2730        2740        2750
TCTTATCTGG TTACATTTTC AGATTGTTAT GAAACTACCC AAATATCCTG CACAATTGCA TGGGATTAAA
AGAATAGACC AATGTAAAAG TCTAACAATA CTTTGATGGG TTTATAGGAC GTGTTAACGT ACCCTAATTT
```

Fig. 42D

```
      2760        2770        2780        2790        2800        2810        2820
TTTTAGTTTC TTTGAAATAG AAGTAGAGTT GTATTGCTGT CACGTCATCA AATAGTTCTG AAGCTATGAA
AAAATCAAAG AAACTTTATC TTCATCTCAA CATAACGACA GTGCAGTAGT TTATCAAGAC TTCGATACTT

      2830        2840        2850        2860        2870        2880        2890
TAAATAAGTT CCGCATTTGT TAGTGATTCT TTGAACATTA GAATTGTTAT GCTTAAGTAG ATAGGGTTAT
ATTTATTCAA GGCGTAAACA ATCACTAAGA AACTTGTAAT CTTAACAATA CGAATTCATC TATCCCAATA

      2900        2910        2920        2930        2940        2950        2960
GTTTGTTTGG AGTTCCCTTA AATCATTTCA TTGCTGACTG CCAGCTGGCA GGAGCATTTG TTGTTGCCTT
CAAACAAACC TCAAGGGAAT TTAGTAAAGT AACGACTGAC GGTCGACCGT CCTCGTAAAC AACAACGGAA

      2970        2980        2990        3000        3010        3020        3030
GACCATGAAT GAAGACCTTC CTGTTCTGAG TGCTCACAAG AAAACATATT TTGATTAATG CACCTTGAAT
CTGGTACTTA CTTCTGGAAG GACAAGACTC ACGAGTGTTC TTTTGTATAA AACTAATTAC GTGGAACTTA

      3040        3050        3060        3070        3080        3090        3100
CCTTAGGATC TTGCAAAGAT GGGCACTTAG CTTTAGAATT GAGTAGTACT TAAATAGCTG TTGTTATCAT
GGAATCCTAG AACGTTTCTA CCCGTGAATC GAAATCTTAA CTCATCATGA ATTTATCGAC AACAATAGTA

      3110        3120        3130        3140        3150        3160        3170
GATTTGTCCT GTAGTGAAAT GTCGACAAAA CAGGAATGCT ACTTTTGACT TCTGATATTT CATGCCTGGC
CTAAACAGGA CATCACTTTA CAGCTGTTTT GTCCTTACGA TGAAAACTGA AGACTATAAA GTACGGACCG

      3180        3190        3200        3210        3220        3230
TTTACTTATG CTCTGTTTGG AACATGGGCA CATATCA GGC AAT GCT ACT CCA GTT CAA AAC ATG
AAATGAATAC GAGACAAACC TTGTACCCGT GTATAGT CCG TTA CGA TGA GGT CAA GTT TTG TAC
                                      G   N   A   T   P   V   Q   N   M>
                                    ___c___c___c__EXON3 ___c___c___c___c___>

   3240        3250        3260        3270        3280        3290
CTA AGT GGC CCA AGT GGG GGA TCG GGC TCA CAG TTG GTA CAG AAT GTT GAT GTC CTT
GAT TCA CCG GGT TCA CCC CCT AGC CCG AGT GTC AAC CAT GTC TTA CAA CTA CAG GAA
 L   S   G   P   S   G   G   S   G   S   Q   L   V   Q   N   V   D   V   L>
___c___c___c___c___c___c___c___c___c__EXON3 ___c___c___c___c___c___c___c___c___>

      3300        3310        3320        3330        3340
GTA AAG CAG CCC ACC AGC TCT TCA TCA AGG GAG CAG TCA GAT GAT GAT GAC ATG AAG
CAT TTC GTC GGG TGG TCG AGA AGT AGT TCC CTC GTC AGT CTA CTA CTA CTG TAC TTC
 V   K   Q   P   T   S   S   S   S   R   E   Q   S   D   D   D   D   M   K>
___c___c___c___c___c___c___c___c___c__EXON3 ___c___c___c___c___c___c___c___c___>

3350        3360        3370        3380        3390        3400
GGA GAA GCT GAG ACC ACT GGA ACT GCA AGA CCT GCT GAT CAA AGA TTA CAA CGA
CCT CTT CGA CTC TGG TGA CCT TGA CGT TCT GGA CGA CTA GTT TCT AAT GTT GCT
 G   E   A   E   T   T   G   T   A   R   P   A   D   Q   R   L   Q   R>
___c___c___c___c___c___c___c___c____EXON3 _c___c___c___c___c___c___c___c___>
```

Fig. 42E

```
    3410        3420        3430        3440        3450        3460        3470
AGGTGATC ATTCATTGCT TCCTTGTAAT ATAGATTCTG TACATAATTA ACCTACCTCG TCATGCATGC
TCCACTAG TAAGTAACGA AGGAACATTA TATCTAAGAC ATGTATTAAT TGGATGGAGC AGTACGTACG

      3480        3490        3500        3510        3520        3530        3540
ATGTGTCCTA TTTTCACCTT AGCCCTTTCA GTTGGATTTC CACTTTCATC CGGTAGCCTT TCAGTTTCCT
TACACAGGAT AAAAGTGGAA TCGGGAAAGT CAACCTAAAG GTGAAAGTAG GCCATCGGAA AGTCAAAGGA

      3550        3560        3570        3580        3590        3600        3610
ATTGCATCGC ATATATGATC TTTTACCTAC CATATTAGTT CTCTGTGTGC CATACTCAGT GCTTAGTGTC
TAACGTAGCG TATATACTAG AAAATGGATG GTATAATCAA GAGACACACG GTATGAGTCA CGAATCACAG

      3620        3630        3640        3650        3660        3670        3680
TCGAGCAAGA GAGGAATTTG TATGGCTATT ACACGTAGCA CTTTGCTCTC TACTTGTTTA TTGACATAAG
AGCTCGTTCT CTCCTTAAAC ATACCGATAA TGTGCATCGT GAAACGAGAG ATGAACAAAT AACTGTATTC

      3690        3700        3710        3720        3730        3740        3750
CAATTTGGGA TGAATTAAAT CTGAGTTCAC ATCATATTCC TTATGTCACA AGTTTCTGAA ACCGATTGTA
GTTAAACCCT ACTTAATTTA GACTCAAGTG TAGTATAAGG AATACAGTGT TCAAAGACTT TGGCTAACAT

      3760        3770        3780        3790        3800        3810        3820
TCTAGTATCT GGTTGATGCA CCCCCATCTT GGATTTGCAA ATCAAAGTTA TACTCCCTAG AGAGCTTTAC
AGATCATAGA CCAACTACGT GGGGGTAGAA CCTAAACGTT TAGTTTCAAT ATGAGGGATC TCTCGAAATG

      3830        3840        3850        3860        3870        3880        3890
CTTTCATAAA GCAATTACCC CAATAAACCA CGGATTTGAT AGCTATTGAC TATGATTACC AGAATTCATT
GAAAGTATTT CGTTAATGGG GTTATTTGGT GCCTAAACTA TCGATAACTG ATACTAATGG TCTTAAGTAA

      3900        3910        3920        3930        3940        3950        3960
TGGCAGCTAT TTTCTCAATT TAAGTTTGGT ATTAGTCTCA GTTGGCTGTA AAATAATGTC ACGGTAGGGT
ACCGTCGATA AAAGAGTTAA ATTCAAACCA TAATCAGAGT CAACCGACAT TTTATTACAG TGCCATCCCA

      3970        3980        3990        4000        4010        4020        4030
ACATGTATGT GCAGCATACA AGGTATGGGT GAGTTATGAT ATGGACAGTG TGTACACCCC ACATTTGCTC
TGTACATACA CGTCGTATGT TCCATACCCA CTCAATACTA TACCTGTCAC ACATGTGGGG TGTAAACGAG

      4040        4050        4060        4070        4080        4090        4100
ACTAAAATCA AAATATTCAA ACGTCACGTG ATGATATGGT GGATTGCATT ATACCTTGTA TTGTTTATTA
TGATTTTAGT TTTATAAGTT TGCAGTGCAC TACTATACCA CCTAACGTAA TATGGAACAT AACAAATAAT

      4110        4120        4130        4140        4150        4160        4170
TGTTACTTGT GCTAGACAAT AATATAGGCT GTTCTTTTGG GTGATTTTGT ATGAAGATGT TGAGCAAGCA
ACAATGAACA CGATCTGTTA TTATATCCGA CAAGAAAACC CACTAAAACA TACTTCTACA ACTCGTTCGT

      4180        4190        4200        4210        4220        4230
CTTCTCGATA TAATGCTAGT TTTGTTGACC TGTTCC AGG AAG CAA TCC AAT CGG GAG TCA GCC
GAAGAGCTAT ATTACGATCA AAACAACTGG ACAAGG TCC TTC GTT AGG TTA GCC CTC AGT CGG
                                        R   K   Q   S   N   R   E   S   A>
                                    ___d___d___d__EXON4____d___d___d___d___>
```

Fig. 42F

```
      4240          4250          4260          4270          4280          4290
AGG CGC TCA AGA AGC AGA AAG GCA GCT CAC TTG AAT GAG CTG GAG GCA CAG GTGTGA
TCC GCG AGT TCT TCG TCT TTC CGT CGA GTG AAC TTA CTC GAC CTC CGT GTC CACACT
 R   R   S   R   S   R   K   A   A   H   L   N   E   L   E   A   Q>
___d___d___d___d___d___d___d__EXON4____d___d___d___d___d___d___d___>

      4300        4310        4320        4330        4340        4350        4360
TAGTTCACAT AGTTATTTTC GATAAGACAT AAAATCCTAA ATTACTGGCT ACTGACTTCA GTTATGGATT
ATCAAGTGTA TCAATAAAAG CTATTCTGTA TTTTAGGATT TAATGACCGA TGACTGAAGT CAATACCTAA

      4370          4380          4390          4400          4410          4420
TACTTGTTAC AG GTA TCG CAA TTA AGA GTC GAG AAC TCC TCG CTG TTA AGG CGT CTT GCT
ATGAACAATG TC CAT AGC GTT AAT TCT CAG CTC TTG AGG AGC GAC AAT TCC GCA GAA CGA
               V   S   Q   L   R   V   E   N   S   S   L   L   R   R   L   A>
             ___e___e___e___e___e___e____EXON5__e___e___e___e___e___e___e___e___>


          4430          4440          4450          4460          4470
GAT GTT AAC CAG AAG TAC AAT GAT GCT GCT GTT GAC AAT AGA GTG CTA AAA GCA GAT
CTA CAA TTG GTC TTC ATG TTA CTA CGA CGA CAA CTG TTA TCT CAC GAT TTT CGT CTA
 D   V   N   Q   K   Y   N   D   A   A   V   D   N   R   V   L   K   A   D>
___e___e___e___e___e___e___e___e___e__EXON5____e___e___e___e___e___e___e___e___>

4480          4490          4500          4510          4520          4530          4540
GTT GAG ACC TTG AGA GCA AAG GT ATGCTATATA TGCCTTTTGC AATATGCATC CCATGGATTG
CAA CTC TGG AAC TCT CGT TTC CA TACGATATAT ACGGAAAACG TTATACGTAG GGTACCTAAC
 V   E   T   L   R   A   K>
___e___e__EXON5____e___e___>

      4550          4560          4570          4580          4590          4600          4610
CTACTTTGGC TTGTTTCAAA CTTTCAACGT GACTTGTGTA CCCTGTTATT AGAAGAATAA TCCCGCCTAC
GATGAAACCG AACAAAGTTT GAAAGTTGCA CTGAACACAT GGGACAATAA TCTTCTTATT AGGGCGGATG

      4620          4630          4640          4650          4660          4670          4680
CATTATACTC TATAAATCAC CATTTGGCCA GTCCAAACAT GATTATTAAA TCAGGTCAAT CTGAACATTG
GTAATATGAG ATATTTAGTG GTAAACCGGT CAGGTTTGTA CTAATAATTT AGTCCAGTTA GACTTGTAAC

      4690          4700          4710          4720          4730          4740
AAATGTATCA AAAATTCGCA GGTG AAG ATG GCA GAG GAC TCG GTG AAG CGG GTG ACA GGC
TTTACATAGT TTTTAAGCGT CCAC TTC TAC CGT CTC CTG AGC CAC TTC GCC CAC TGT CCG
                            K   M   A   E   D   S   V   K   R   V   T   G>
                          ___f___f___f___f____EXON6__f___f___f___f___f___>

          4750          4760          4770          4780          4790
ATG AAC GCG TTG TTT CCC GCC GCT TCT GAT ATG TCA TCC CTC AGC ATG CCA TTC AAC
TAC TTG CGC AAC AAA GGG CGG CGA AGA CTA TAC AGT AGG GAG TCG TAC GGT AAG TTG
 M   N   A   L   F   P   A   A   S   D   M   S   S   L   S   M   P   F   N>
___f___f___f___f___f___f___f___f___f__EXON6____f___f___f___f___f___f___f___f___>
```

Fig. 42G

```
4800          4810          4820          4830          4840          4850
 AGC TCC CCA TCT GAA GCA ACG TCA GAC GCT GCT GTT CCC ATC CAA GAT GAC CCG AAC
 TCG AGG GGT AGA CTT CGT TGC AGT CTG CGA CGA CAA GGG TAG GTT CTA CTG GGC TTG
  S   S   P   S   E   A   T   S   D   A   A   V   P   I   Q   D   D   P   N>
 ___f___f___f___f___f___f___f___f__EXON6____f___f___f___f___f___f___f___f___>

    4860          4870          4880          4890          4900          4910
 AAT TAC TTC GCT ACT AAC AAC GAC ATC GGA GGT AAC AAC AAC TAC ATG CCC GAC ATA
 TTA ATG AAG CGA TGA TTG TTG CTG TAG CCT CCA TTG TTG TTG ATG TAC GGG CTG TAT
  N   Y   F   A   T   N   N   D   I   G   G   N   N   N   Y   M   P   D   I>
 ___f___f___f___f___f___f___f___f__EXON6____f___f___f___f___f___f___f___f___>


        4920          4930          4940          4950          4960
 CCT TCT TCG GCT CAG GAG GAC GAG GAC TTC GTC AAT GGC GCT CTG GCT GCC GGC AAG
 GGA AGA AGC CGA GTC CTC CTG CTC CTG AAG CAG TTA CCG CGA GAC CGA CGG CCG TTC
  P   S   S   A   Q   E   D   E   D   F   V   N   G   A   L   A   A   G   K>
 ___f___f___f___f___f___f___f___f__EXON6____f___f___f___f___f___f___f___f___>

          4970          4980          4990          5000          5010
5020
 ATT GGC CGG CCA GCC TCG CTG CAG CGG GTG GCG AGC CTG GAG CAT CTC CAG AAG AGG
 TAA CCG GCC GGT CGG AGC GAC GTC GCC CAC CGC TCG GAC CTC GTA GAG GTC TTC TCC
  I   G   R   P   A   S   L   Q   R   V   A   S   L   E   H   L   Q   K   R>
 ___f___f___f___f___f___f___f___f__EXON6____f___f___f___f___f___f___f___f___>

   5030          5040          5050          5060          5070          5080
 ATG TGC GGT GGG CCG GCT TCG TCT GGG TCG ACG TCC TGA GACCGA AACCCAGAGC
 TAC ACG CCA CCC GGC CGA AGC AGA CCC AGC TGC AGG ACT CTGGCT TTGGGTCTCG
  M   C   G   G   P   A   S   S   G   S   T   S   *>
 ___f___f___f___f___f__EXON6____f___f___f___f___f___>

        5090          5100          5110          5120          5130          5140          5150
 TGCTTCGGTT CTGAAAGACA CTGCGAGCAG GAAATGATGA TTGGACAGGC GTAGACATTG CTAATGCTGT
 ACGAAGCCAA GACTTTCTGT GACGCTCGTC CTTTACTACT AACCTGTCCG CATCTGTAAC GATTACGACA

        5160          5170          5180          5190          5200          5210          5220
 GAGGTTGATG ATTGTTGGTC GTCGTCGTCG TCATTGTGCA TTCTTTGTAA GGGACACCTC TTAGTACCCT
 CTCCAACTAC TAACAACCAG CAGCAGCAGC AGTAACACGT AAGAAACATT CCCTGTGGAG AATCATGGGA

        5230          5240          5250          5260          5270          5280          5290
 CTTCTTCTAA GGGACTTAGT ACCCCTTGTG GATCTCATCG TCCTAAATAC TATACACATT AGCCAAATGT
 GAAGAAGATT CCCTGAATCA TGGGGAACAC CTAGAGTAGC AGGATTTATG ATATGTGTAA TCGGTTTACA


>terminator

                                                                              |
        5300          5310          5320          5330          5340          5350          5360
 TCATTGGTGT GATGGCGTCG TCCCTAATTT GAACGACTGA TTTCAGGCAG CTGCTATGCT ATCATTCAAT
 AGTAACCACA CTACCGCAGC AGGGATTAAA CTTGCTGACT AAAGTCCGTC GACGATACGA TAGTAAGTTA
```

Fig. 42H

```
      5370        5380        5390        5400        5410        5420        5430
AATATTTTGA TCGATGCTTC CTCTTGTCTT TTGCTCTTAA GCAACCAAGC ATAAAGATAT CACTACCTTT
TTATAAAACT AGCTACGAAG GAGAACAGAA AACGAGAATT CGTTGGTTCG TATTTCTATA GTGATGGAAA

      5440        5450        5460        5470        5480        5490        5500
TGAGCTGTTC ATTTGAAGTG CAAAGCTAAG CTCAATATCT CAGGTGTTCA TTTGAAGTTT AAAGGTGAAC
ACTCGACAAG TAAACTTCAC GTTTCGATTC GAGTTATAGA GTCCACAAGT AAACTTCAAA TTTCCACTTG

      5510        5520        5530        5540        5550        5560        5570
TGATAACAAA CGTCAGGCTA TGGTGAATGA AGGGACGTGT ACATCCCTAA TACATGTCAT TTTCATAATC
ACTATTGTTT GCAGTCCGAT ACCACTTACT TCCCTGCACA TGTAGGGATT ATGTACAGTA AAAGTATTAG

      5580        5590        5600        5610        5620        5630        5640
AAATTAGTTG ATGCATTTTC ACCCAGAATC CCATCACAGT TCATCATACA AGCAAGTGTA GTTATTAATG
TTTAATCAAC TACGTAAAAG TGGGTCTTAG GGTAGTGTCA AGTAGTATGT TCGTTCACAT CAATAATTAC

      5650        5660        5670        5680        5690        5700        5710
GTAAATTTTT CGTTTAGAGA AAAAAAAAGG AAGCCTTATA TAAGATTCAC CGGTGGGGTG TGAACAATAA
CATTTAAAAA GCAAATCTCT TTTTTTTTCC TTCGGAATAT ATTCTAAGTG GCCACCCCAC ACTTGTTATT

      5720        5730        5740        5750        5760        5770        5780
TCAATGAATG AGATCGCATC CCGTAAGGGC AGCCTAGCTA GACAAAAATG CATAAAACTC CGTATACCAA
AGTTACTTAC TCTAGCGTAG GGCATTCCCG TCGGATCGAT CTGTTTTTAC GTATTTTGAG GCATATGGTT

      5790        5800        5810        5820        5830        5840        5850
CCACAACAAC GCTTGCGCAC GCGCTCAAAT GGCAGCGACT TCATCGCTTT CGCGGGCAAG AAACGAATCA
GGTGTTGTTG CGAACGCGTG CGCGAGTTTA CCGTCGCTGA AGTAGCGAAA GCGCCCGTTC TTTGCTTAGT

      5860        5870        5880        5890        5900        5910        5920
AGTGATACAT TGGCAGGGAA CCACCAAAAG AAGGCCATCC AATCCAATCC ACTCCAACGC GGCATGGAAG
TCACTATGTA ACCGTCCCTT GGTGGTTTTC TTCCGGTAGG TTAGGTTAGG TGAGGTTGCG CCGTACCTTC

      5930        5940        5950        5960        5970        5980        5990
ACAAGACAGA TGATTCACAG CTATCTTCTG CTTCTACAAG TTTGATACTT TGTACTGTCC TTTCAGGGAA
TGTTCTGTCT ACTAAGTGTC GATAGAAGAC GAAGATGTTC AAACTATGAA ACATGACAGG AAAGTCCCTT

      6000        6010        6020        6030        6040        6050        6060
AAAAGAGCAT CAGATTAGTC TGATCTCGGG CGCGTTGAGT TCTTGTGGGA GATCTTGTTG TGGAGTGGCA
TTTTCTCGTA GTCTAATCAG ACTAGAGCCC GCGCAACTCA AGAACACCCT CTAGAACAAC ACCTCACCGT

      6070        6080        6090        6100        6110        6120        6130
GGAGTGACGA TCGGCTGCCC CGTTTTCTTC TACCGAAACA TCGCCAGTAA AGAAGCCAAA AAGACAATAA
CCTCACTGCT AGCCGACGGG GCAAAAGAAG ATGGCTTTGT AGCGGTCATT TCTTCGGTTT TTCTGTTATT

      6140        6150        6160        6170        6180        6190        6200
TACGGCAATG GGGATCGCCC ATCTGCATAA AACATTGCAT GACGGAACTG ATTAATACAA GAATGACATG
ATGCCGTTAC CCCTAGCGGG TAGACGTATT TTGTAACGTA CTGCCTTGAC TAATTATGTT CTTACTGTAC

      6210        6220
TAAGCTGATA ATTACGCGTG CAAGCTT
ATTCGACTAT TAATGCGCAC GTTCGAA
```

Fig. 42I

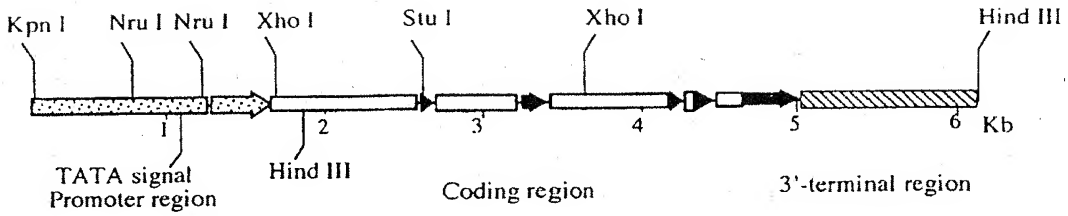

Kpn I
Nru I
Nru I
Xho I
Stu I
Xho I
Hind III
1
2
3'
4
5
6
Kb
TATA signal
Promoter region
Hind III
Coding region
3'-terminal region

Fig. 43

```
        10         20         30         40         50         60         70         80         90
AAGCTTGCAT GCCTGCAGGG AGGAGAGGGG AGAGATGGTG AGAGAGGAGG AAGAAGAGGA GGGGTGACAA TGATATGTGG GCCATGTGGC
                                                                  >Reb_site1

       100        110        120        130        140        150        160        170        180
CCCCACCATT TTTTAATTCA TTCTTTTGTT GAAACTGACA TGTGGGTCCC ATGAGAATTA TTATTTTTCG GATCGAATT GCCACGTAAGC

                                   >Reb_site1  >reb_site2

       190        200        210        220        230        240        250        260        270
GCTACGTCAA TGCTACGTCA GATGAAGACC GAGTCAAATT AGCCACGTAAG CGCCACGTCAG CCAAAACC ACCATCCAAA CCGCCGAGGG

       280        290        300        310        320        330        340        350        360
ACCTCATCTG CACTGGTTTT GATAGTTGAG GGACCCGTTG TATCTGGTTT TTCGATTGAA GGACGAAAAT CAAATTTGTT GACAAGTTAA

       370        380        390        400        410        420        430        440        450
GGGACCTTAA ATGAACTTAT TCCATTTCAA AATATTCTGT GAGCCATATA TCCGTGGGCT TCCAATCCTC CTCAAATTAA AGGGCCTTTT

       460        470        480        490        500        510        520        530        540
TAAAATAGAT AATTGCCTTC TTTCAGTCAC CCATAAAAGT ACAAAACTAC TACCAACAAG CAACATGCGC AGTTACACAC ATTTTCTGCA

       550        560        570        580        590        600        610        620        630
CATTTCCACC ACGTCACAAA GAGCTAAGAG TTATCCCTAG GACAATCTCA TTAGTGTAGA TACATCCATT AATCTTTTAT CAGAGGCAAA

       640        650        660        670        680        690        700        710        720
CGTAAAGCCG CTCTTTATGA CAAAAATAGG TGACACAAAA GTGTTATCTG CCACATACAT AACTTCAGAA ATTACCCAAC ACCAAGAGAA

       730        740        750        760        770        780        790        800        810
AAATAAAAAA AAATCTTTTT GCAAGCTCCA AATCTTGGAA ACCTTTTTCA CTCTTTGCAG CATTGTACTC TTGCTCTTTT TCCAACCGAT

       820        830        840        850        860        870        880        890        900
CCATGTCACC CTCAAGCTTC TACTTGATCT ACACGAAGCT CACCGTGCAC ACAACCATGG CCACAAAAAC CCTATAAAAC CCCATCCGAT

       910        920        930        940        950        960        970        980        990
CGCCATCATC TCATCATCAG TTCATCACCA ACAAACAAAA GAGGAAAAAA AACATATACA CTTCTAGTGA TTGTCTGATT GATCATCAAT

      1000       1010       1020       1030       1040       1050       1060
CTAGAGGATC CCCGGGTGGT CAGTCCCTT ATG
                          Gus Start site
```

Fig. 44

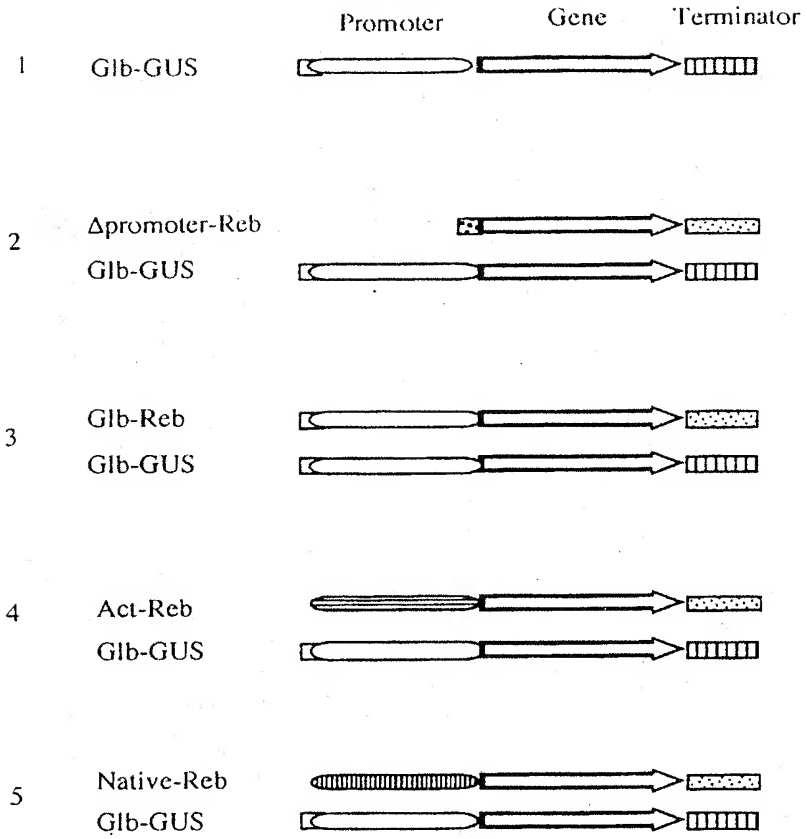


Fig. 45A

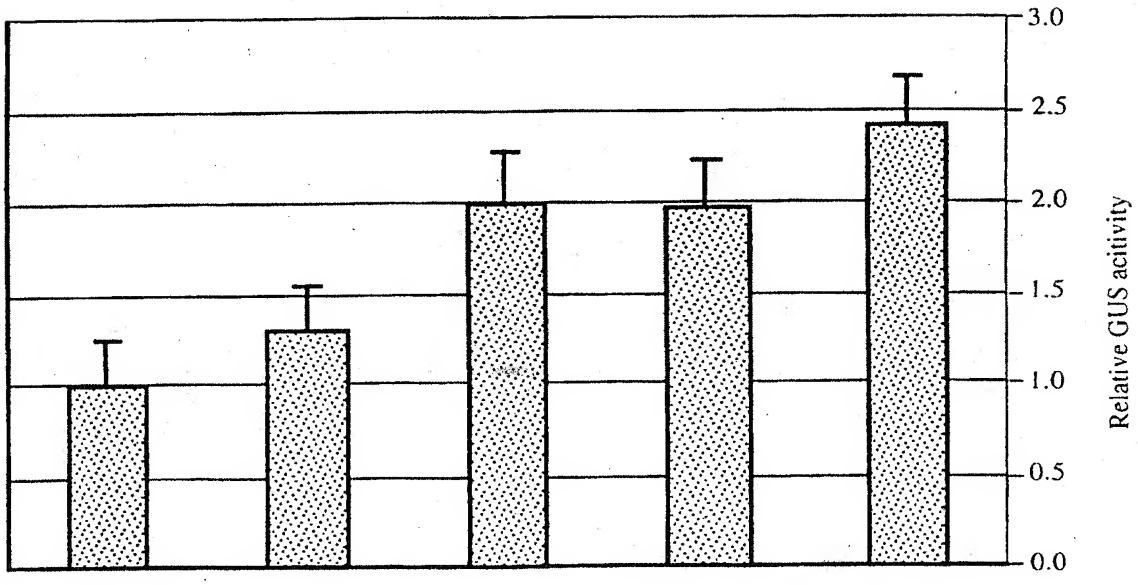


Fig. 45B

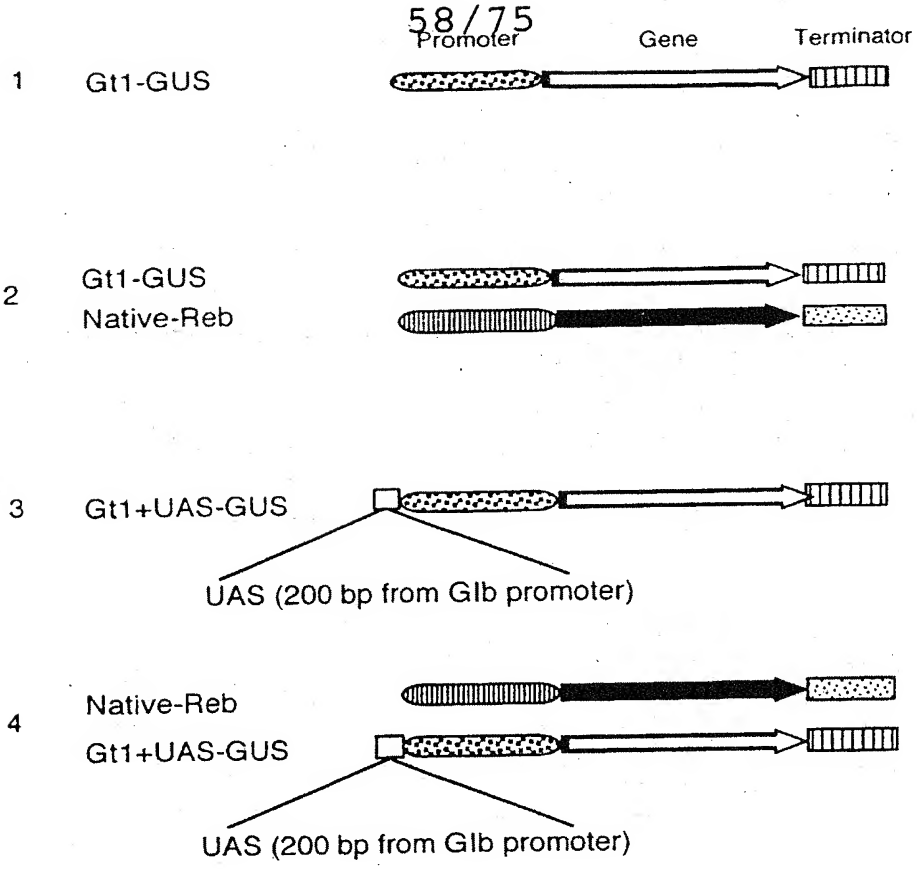


Fig. 46A

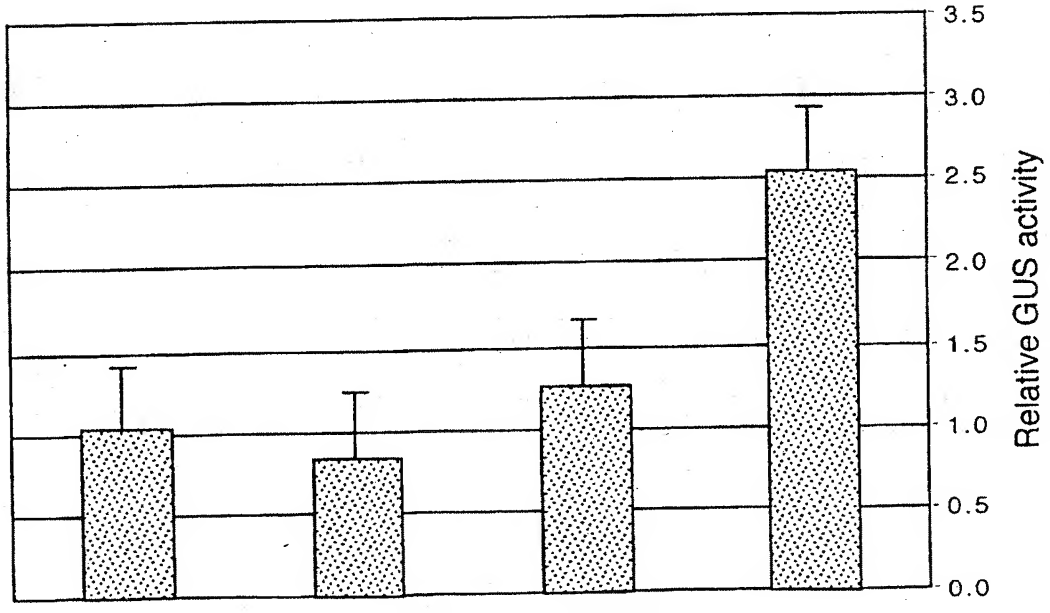


Fig. 46B

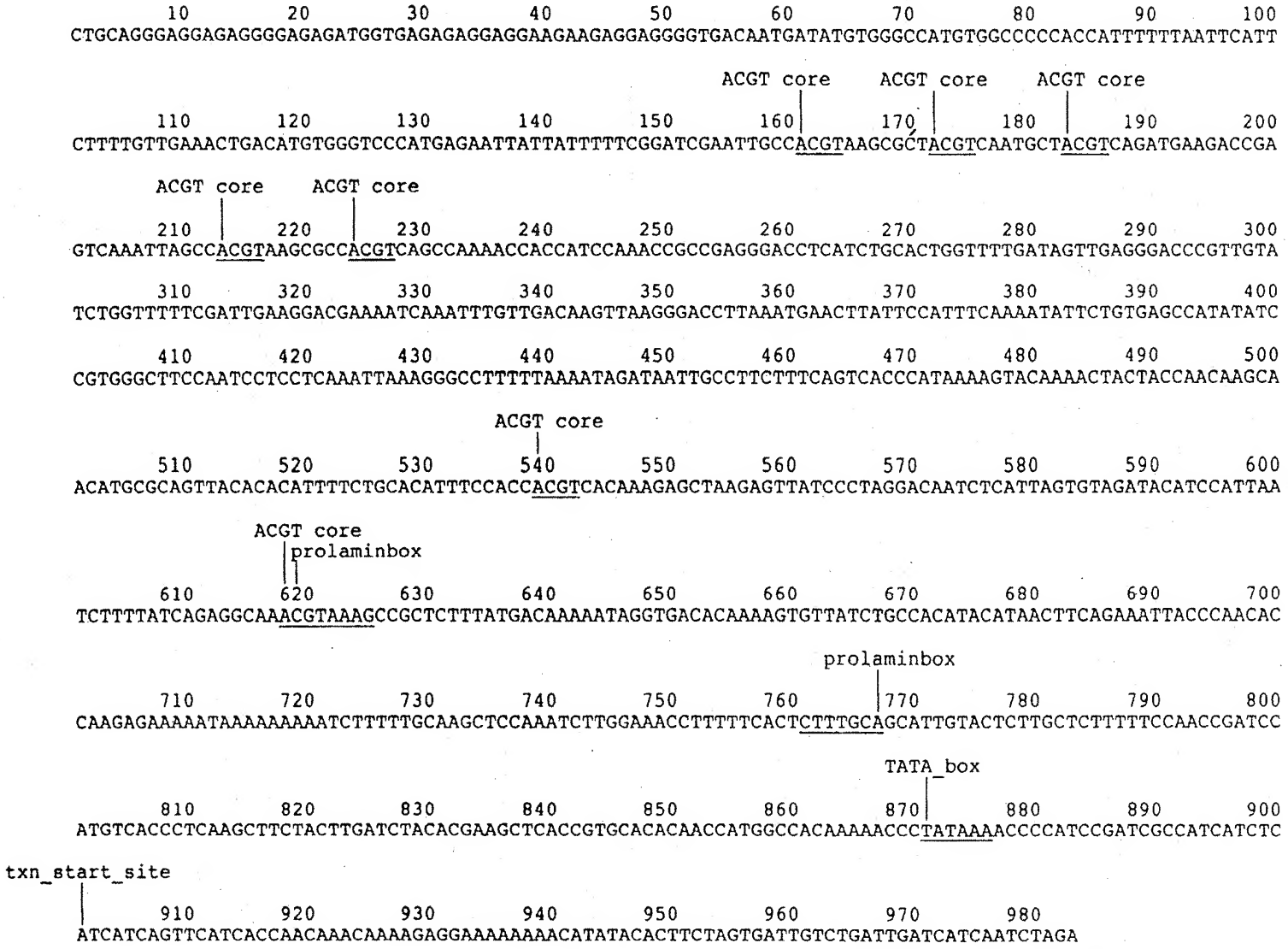


Fig. 47

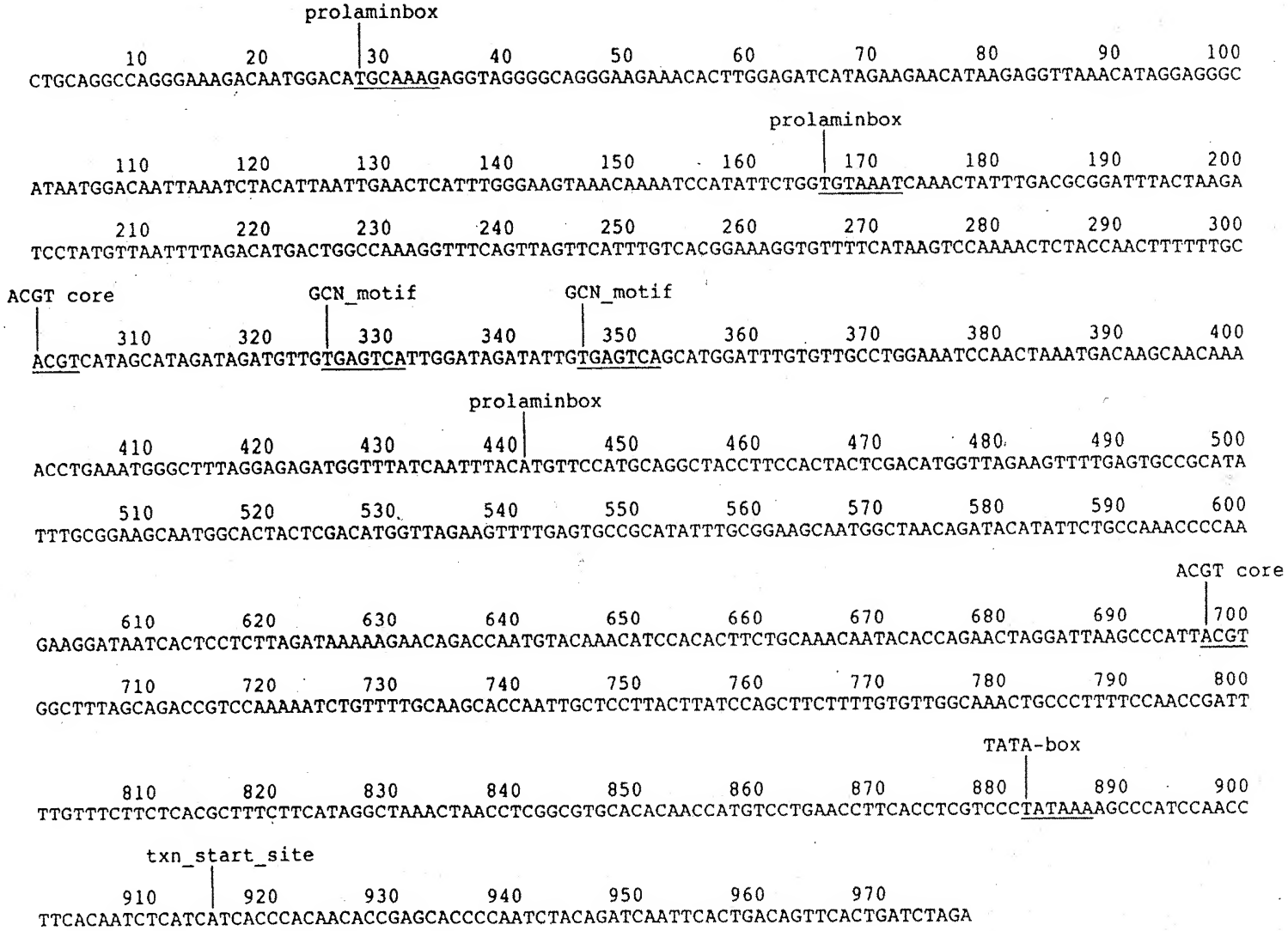

prolaminbox
10 20 30 40 50 60 70 80 90 100
CTGCAGGCCAGGGAAAGACAATGGACATGCAAAGAGGTAGGGGCAGGGAAGAAACACTTGGAGATCATAGAAGAACATAAGAGGTTAAACATAGGAGGGC
prolaminbox
110 120 130 140 150 160 170 180 190 200
ATAATGGACAATTAAATCTACATTAATTGAACTCATTTGGGAAGTAAACAAAATCCATATTCTGGTGTAAATCAAACTATTTGACGCGGATTTACTAAGA
210 220 230 240 250 260 270 280 290 300
TCCTATGTTAATTTTAGACATGACTGGCCAAAGGTTTCAGTTAGTTCATTTGTCACGGAAAGGTGTTTTCATAAGTCCAAAACTCTACCAACTTTTTTGC
ACGT core
GCN_motif
GCN_motif
310 320 330 340 350 360 370 380 390 400
ACGTCATAGCATAGATAGATGTTGTGAGTCATTGGATAGATATTGTGAGTCAGCATGGATTTGTGTTGCCTGGAAATCCAACTAAATGACAAGCAACAAA
prolaminbox
410 420 430 440 450 460 470 480 490 500
ACCTGAAATGGGCTTTAGGAGAGATGGTTTATCAATTTACATGTTCCATGCAGGCTACCTTCCACTACTCGACATGGTTAGAAGTTTTGAGTGCCGCATA
510 520 530 540 550 560 570 580 590 600
TTTGCGGAAGCAATGGCACTACTCGACATGGTTAGAAGTTTTGAGTGCCGCATATTTGCGGAAGCAATGGCTAACAGATACATATTCTGCCAAACCCCAA
ACGT core
610 620 630 640 650 660 670 680 690 700
GAAGGATAATCACTCCTCTTAGATAAAAAGAACAGACCAATGTACAAACATCCACACTTCTGCAAACAATACACCAGAACTAGGATTAAGCCCATTACGT
710 720 730 740 750 760 770 780 790 800
GGCTTTAGCAGACCGTCCAAAAATCTGTTTTGCAAGCACCAATTGCTCCTTACTTATCCAGCTTCTTTTGTGTTGGCAAACTGCCCTTTTCCAACCGATT
TATA-box
810 820 830 840 850 860 870 880 890 900
TTGTTTCTTCTCACGCTTTCTTCATAGGCTAAACTAACCTCGGCGTGCACACAACCATGTCCTGAACCTTCACCTCGTCCCTATAAAAGCCCATCCAACC
txn_start_site
910 920 930 940 950 960 970
TTCACAATCTCATCATCACCCACAACACCGAGCACCCCAATCTACAGATCAATTCACTGACAGTTCACTGATCTAGA

Fig. 48

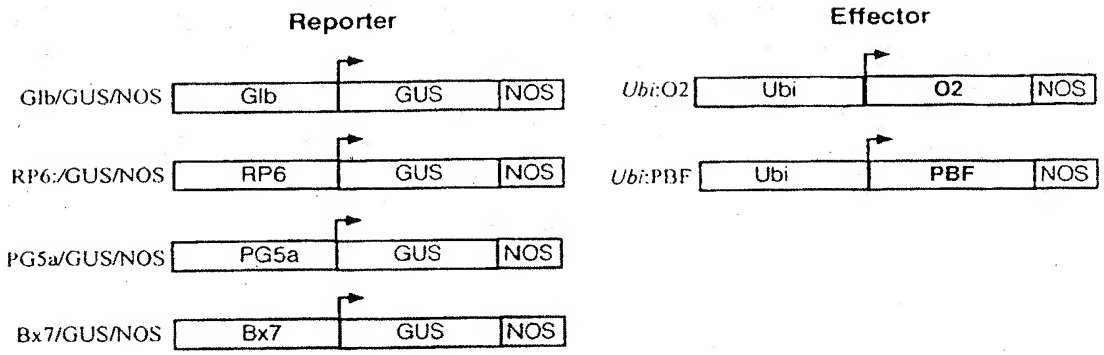


Fig. 49A

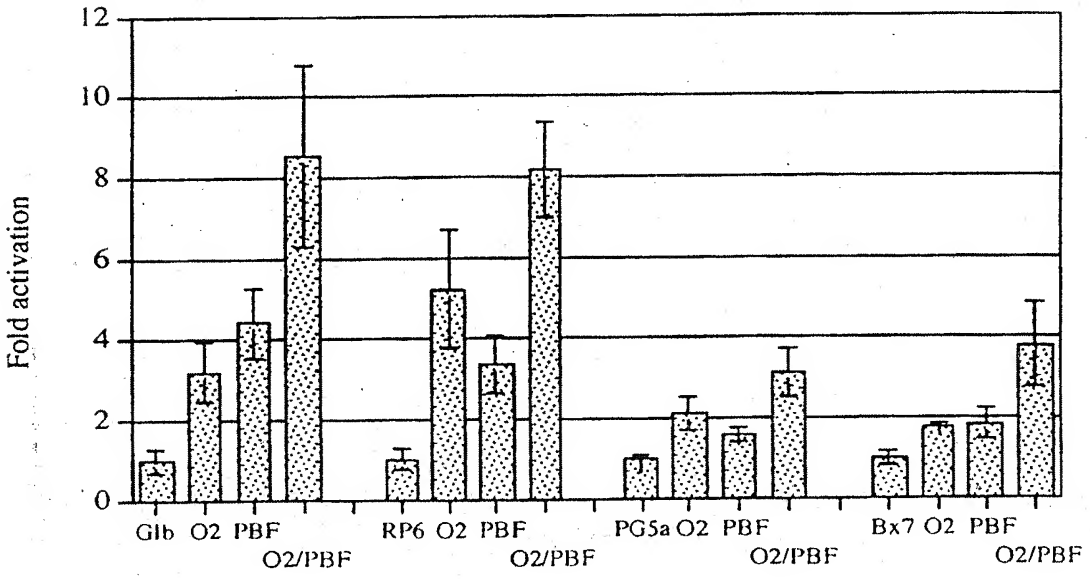


Fig. 49B

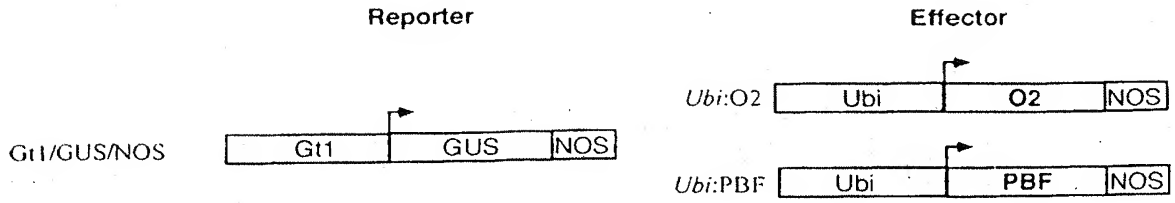


Fig. 50A

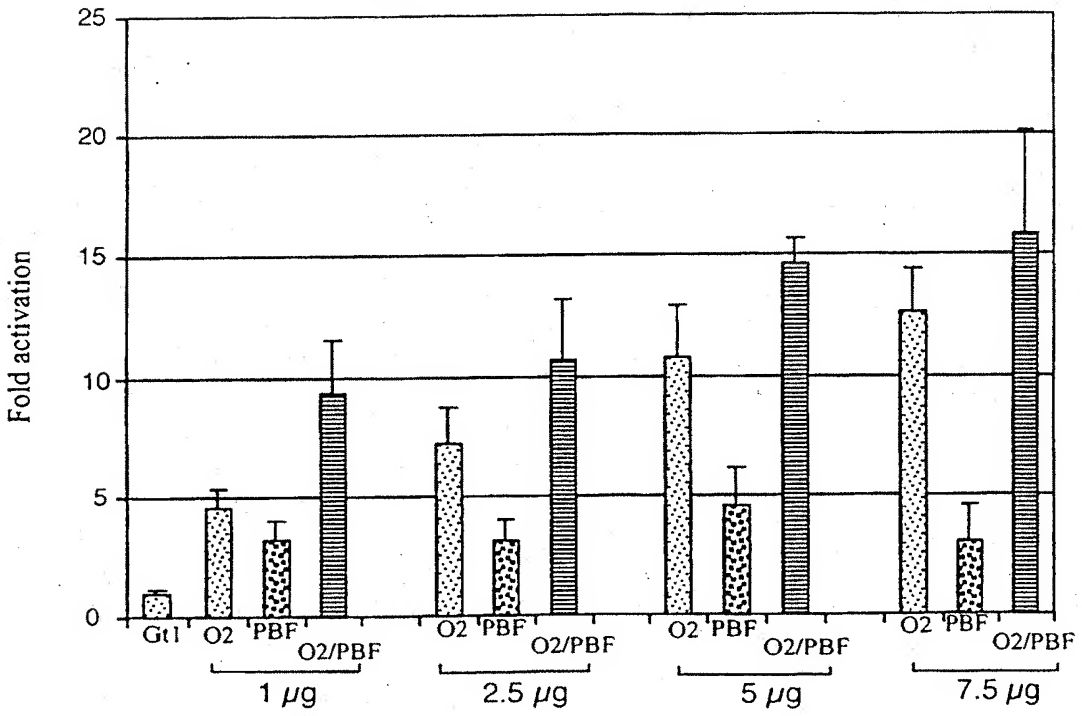


Fig. 50B

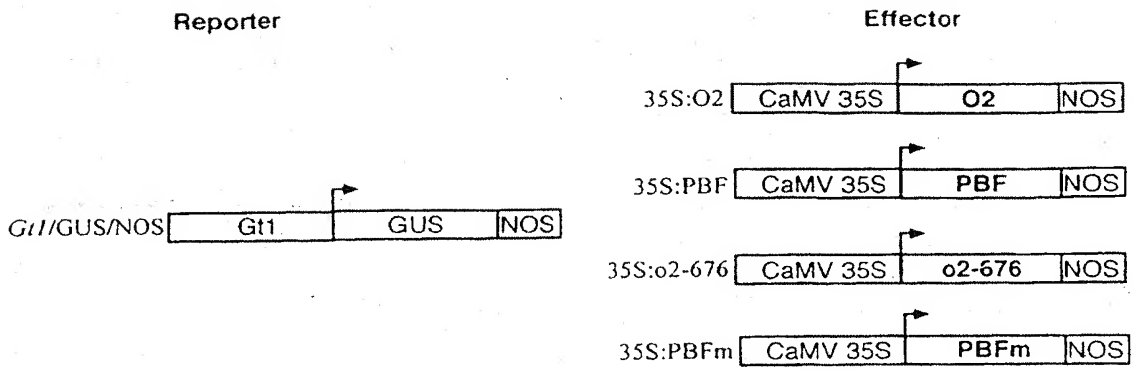


Fig. 51A

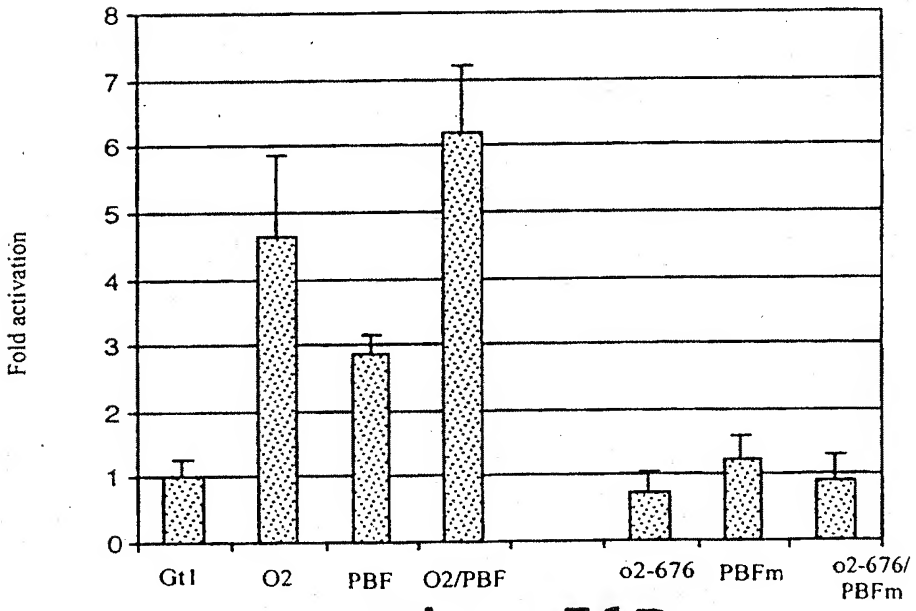


Fig. 51B

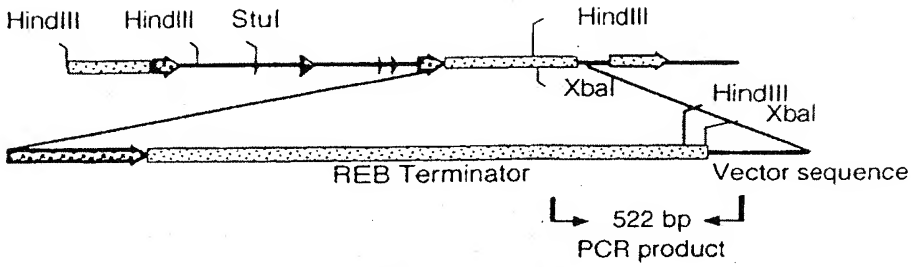


Fig. 52A

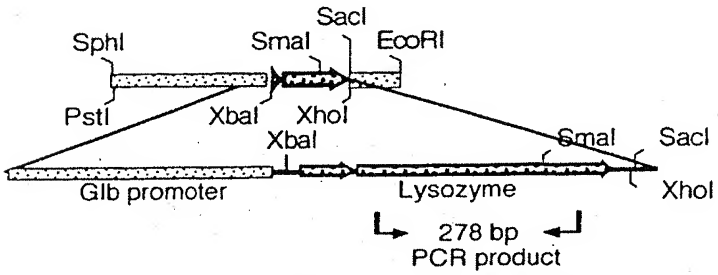


Fig. 52B

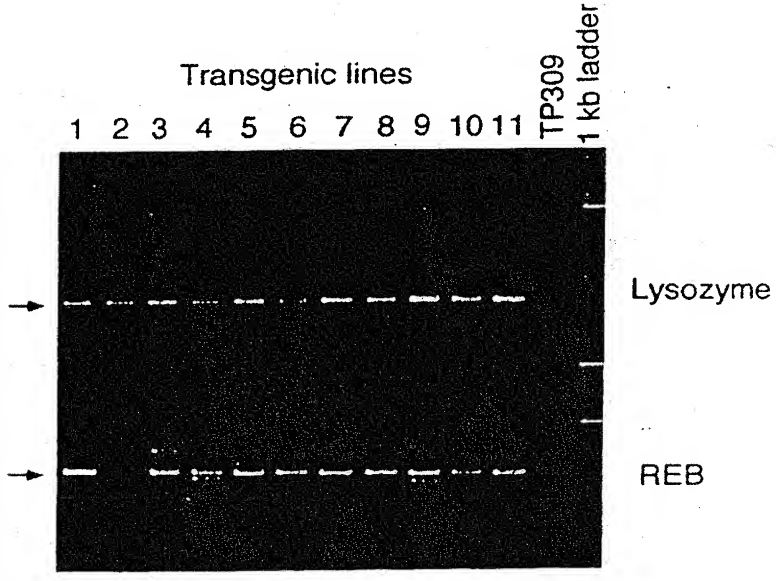


Fig. 52C

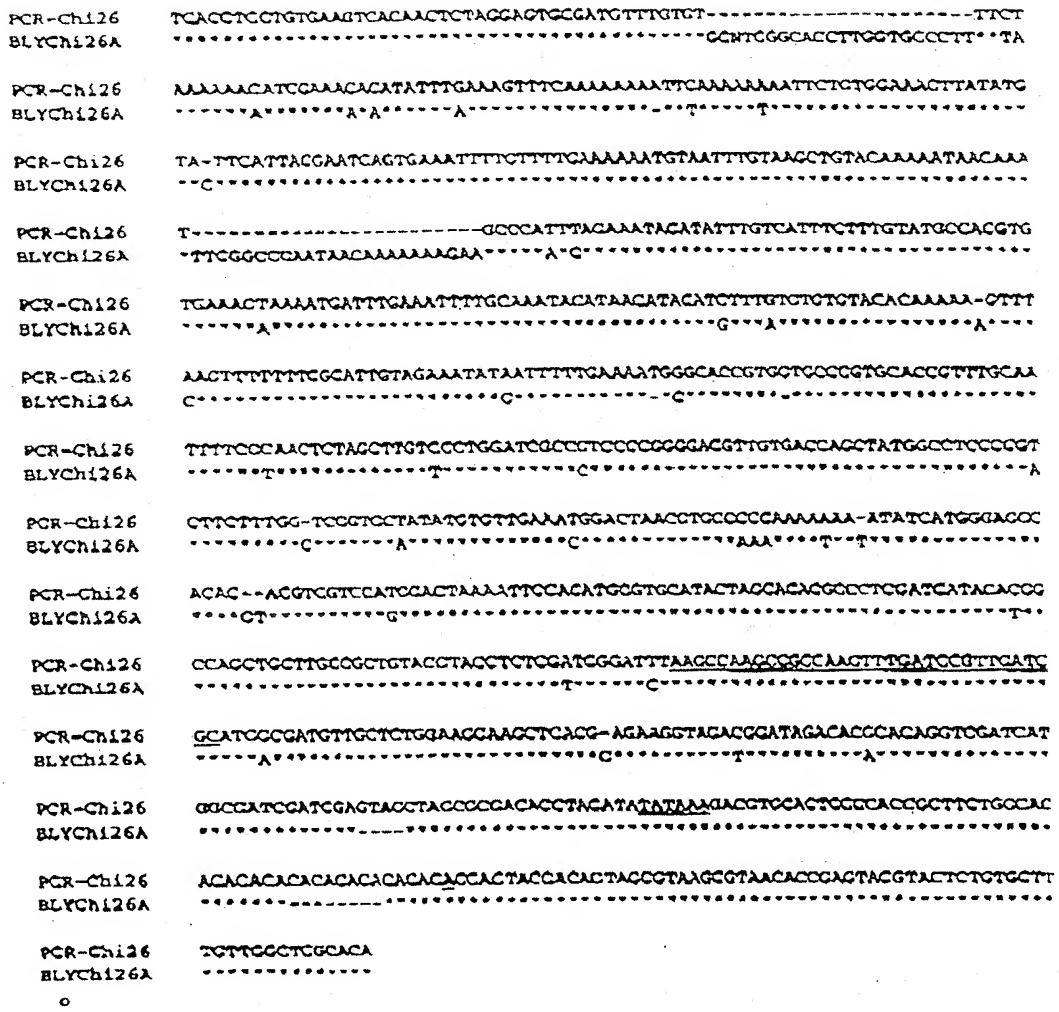

PCR-Chi26
BLYChi26A
TCACCTCCTGTGAAGTCACAACTCTAGGAGTGCGATGTTTGTGT----------TTCT
GCNTCGGCACCTTGGTGCCCTT**TA
AAAAAACATCGAAACACATATTTGAAAGTTTCAAAAAAAAATTCAAAAAAAAATTCTGTGGAAACTTATATG
TA-TTCATTACGAATCAGTGAAATTTTGTTTTGAAAAAATGTAATTTGTAAGCTGTACAAAAATAACAAA
T----------GCCCATTTAGAAATACATATTTGTCATTTCTTTGTATGCCACGTG
-TTCGGGCCGAATAACAAAAAAAGAA*****A*G
TGAAACTAAAATGATTTGAAATTTTGCAAATACATAACATACATCTTTGTGTGTGTACACAAAAA-GTTT
AACTTTTTTTTCGCATTGTAGAAATATAATTTTTGAAAATGGGCACCGTGGTGCCCGTGCACCGTTTGCAA
TTTTCCCAACTCTAGCTTGTCCCTGGATCGCCGTCCCCGGGGACGTTGTGACCAGCTATGGCCTCCCCGT
CTTCTTTGG-TCCGTGCTATATGTGTTGAAATGGACTAACCTGCCCCCAAAAAAA-ATATCATGGGACCC
ACAC--ACGTCGTCCATCCACTAAAATTCCACATGCGTGCATACTACCAGCAGGCCCTCGATCATACACCG
CCAGCTGCTTGCCGCTGTACCTACCTCTCGATCGGATTTAAGCCAAGCCGCCAAGTTTGATCCGTTGATC
GCATCGCCGATGTTGCTCTGGAAGCAAGCTCACG-AGAAGGTAGACGGATAGACACCCACAGGTCGATCAT
GGCGATCGATCGAGTACCTACCCCGACACCTAGATATATAAAGACGTCCAGTCGCCACCGCTTCTGCCAC
ACACACACACACACACACACCACTACCACACTACCGTAAGCGTAACACCGAGTACGTACTCTGTGCTT
TGTTGGCTCGCACA

Fig. 53A

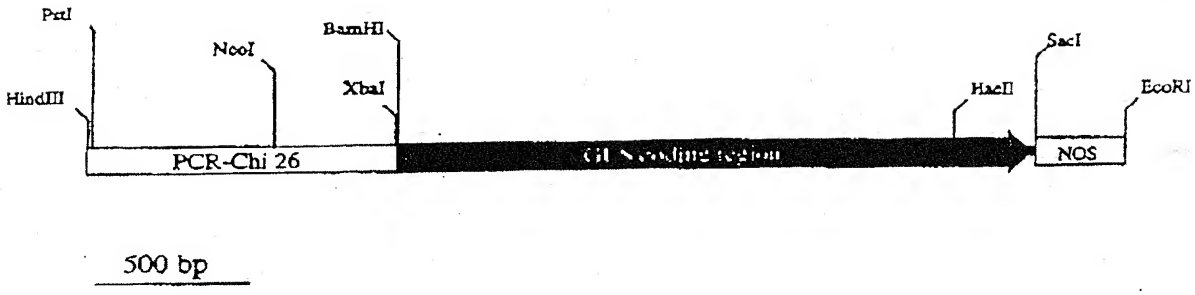


Fig. 53B

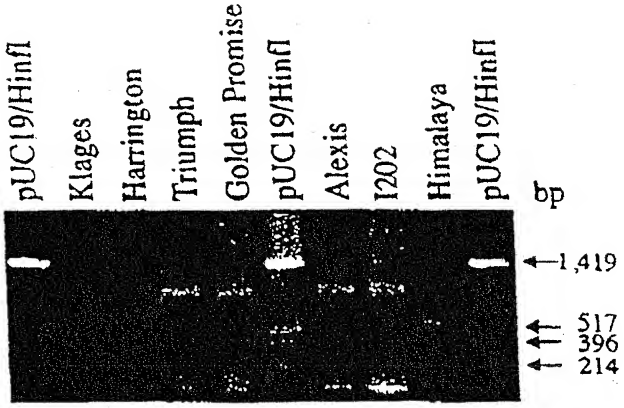

pUC19/HinfI
Klages
Harrington
Triumph
Golden Promise
pUC19/HinfI
Alexis
I202
Himalaya
pUC19/HinfI
bp
1,419
517
396
214

Fig. 54A

```
PCR-Lptl    TTCGAACAATCTCGACGACTTAATTGGAGACTCCCAACAACACCACGAACGTTCATCATAACGAAATATC
BLYLptl     ************************************************************

PCR-Lptl    GCTTCGAGGTAACCTCAAATCCTCCGCCCAATTTTTACAACCTAATTCAAGACCTCGACGCTTGCGTGGA
BLYLptl     ************************************************************

PCR-Lptl    GCTTTACACTATAATGATTGAGCTCCAACG-CATCACCAAGCTTCTATCACCCCAAAACATCCAAGAAAG
BLYLptl     ***************************G************************************

PCR-Lptl    ATATGTACTAGGATACCAAGCACCCAACAGTAAACCGCAGGAAGTATAATATAAGGCCCTGTTTGATAACA
BLYLptl     ************************************************************

PCR-Lptl    AAGTAGTAAAAAAAACTAAAGTATTAAAAACTCCAGTAATTTTACG-GTACATAGAAAATACCATCGTTTT
BLYLptl     ********************************************T***********************

PCR-Lptl    AATATAATAATATTTTTTCCAGTATTCACAATGTAGAGAAACTGTTTCATTACGCCACATATTACTGCAG
BLYLptl     ************************************************************

PCR-Lptl    TTTAGATCGAGCAAGTACACGCCAAGAAGATAACGACGTCCCACCCCTTCTTTTGGCCTTCTCTGTTTTT
BLYLptl     ************************************************************

PCR-Lptl    TAAAAAGAGGTCTGGGGTTAGTTTTTTCAATACTGCAGTTTTAAAATCACAGTTCTTAGACGCAACC-AA
BLYLptl     ********************************************A***************G--

PCR-Lptl    ACACCTCATTGTAAATAAAACTATGATAATCTCCAAAACTGCAGTATTCTAAAAATACTACAAAAATTCT
BLYLptl     ************************************************************

PCR-Lptl    TTGTTATCAAACAGGGCCTAAGGAGTTAAAAAAATTTAGCCGTAACTGAGACTCGGCGAGGCACCAGCAG
BLYLptl     ************************************************************

PCR-Lptl    CTAGCAGTCATCAACACTTGATGGTTGGCAAAGGCGACTCGACGTGTCGCGGGGCTCGGCCTGAGCCGGA
BLYLptl     ************************************************************

PCR-Lptl    GATACAATCTGTTCTCCAGTAACCCCGTCGATTTGGCCCCCCCGACTAAAGCATCCACGCATCTCTCGCTC
BLYLptl     ************************************************************

PCR-Lptl    GAACCCCTATTTAAGCCCCTCCATTCCTCCCAACATTCTCGACACCTCCACGAGTTGCTCATCACTAGCT
BLYLptl     ***********************************************************,,,,,,

PCR-Lptl    AGTACGTTCTACTGTTAGCTACAGATTAAGAAGTGATC
BLYLptl     ***************************************
```

Fig. 54B

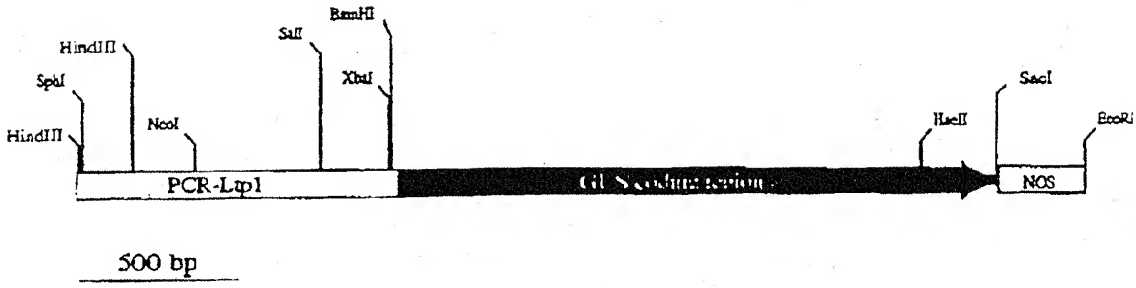


Fig. 54C

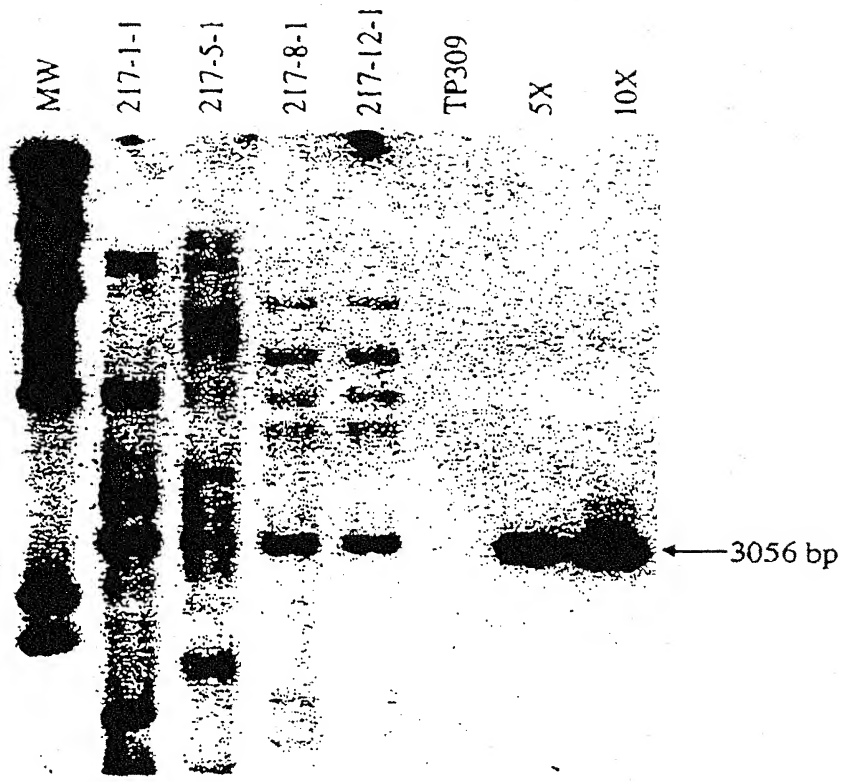


Fig. 55A

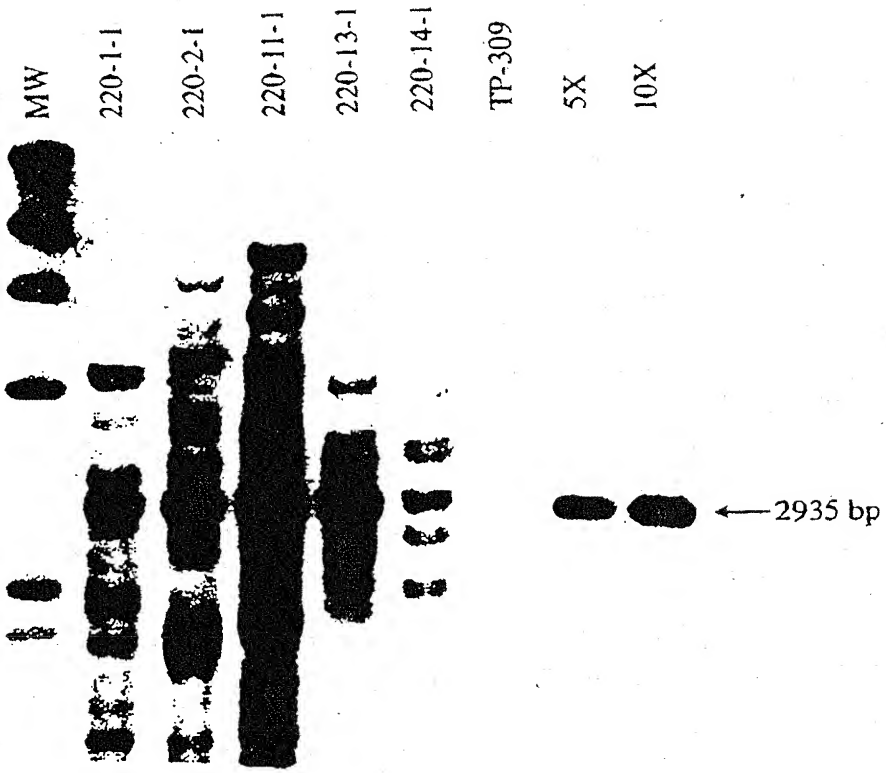


Fig. 55B

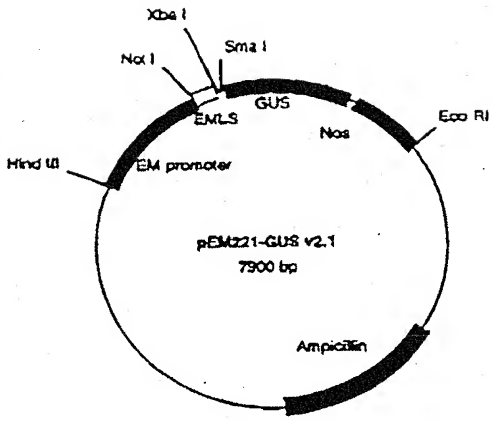


Fig. 56

```
   1 GAGCCTCCTAGGCTA GAAGAAATCCTAACG CCACCACTAATCTGA AGCACGAAACAAATC AATTAAAGAAGCCAT
  76 AAATGATAAATGAAT AATCCTTCATGCATA AAAATATTTAGTTAT CCATAAAGAGACGAC AGACGTATGCAACGT
 151 TCATATTAATTGTCA CTCTACAGTGTCACC GAGTTGTCGGTCCAA GCAATTTTACCCATA CATGAAAACCATCTG
 226 AGAAAGGGCATGAGA GTAGTACTGCGTCAA TACACATGCAGTTGC ACACAGTAGCTGAGA GAGCGAGCCGGAAAC
 301 ATCTGGTCGCCGGTC CACGATCGGTCTACG CATGGGCTTCCCCCC GACCAGGTGGTGCGT GCGAGCCTACCACAT
 376 GTCCACATGCATGCA CGGTGATCGACATCC CGATCGACGTGTCGA CAGTGCAGCCACCCC GCCGCTCGCTCGCTC
 451 GCGCTCGCGAGTCGC GCTAGCACGCACATA CGGAGTAGTATATGC GTATGCATAGGCACG CGCGCGCACGCGCGA
 526 TGCAGCGAACCAGAG CGCGCCGCGCACGGG CGGAGGCATAGCTTG CGCACGTACGTGTCG CGCTCCGGCGGCCTC
 601 GCCACGACGCGTGTC GCGGAGGCGCGCGCG CGCACCGAGCCGTCC GCGTCGCCGCGCGAC ACGTACGTTGCCGGC
 675 GCCCTGATCCCGCCG CGGAGTCGCCGCCTA TAAGAAGGACAGCGC GGCCGCGTCGTCCAT CCCCATCAGCTCAAG
 751 CCGCAGAAGACATAC ACACACAAACACAAG CCACCCTTCCGATTT GTTCATCGATCAGTT CGCAGCGTACGTCAG
 826 GCTAGCTAACTAGTG TTTGGCA ATG GCG TCC GGG CAG CAG CAG CAG GGC AGG TCG GAG CTG GAC
                           M   A   S   G   Q   Q   Q   Q   G   R   S   E   L   D
 900 CGC ATG GCC AGG GAG GGC CAG ACC GTC GTC CCC GGC GGC ACC GGC GGC AAG AGC CTC GAG
       R   M   A   R   E   G   Q   T   V   V   P   G   G   T   G   G   K   S   L   E
 950 GCC CAG GAG AAC CTC GCC GAG G | GTATGCAAAATAATA CTGAAACTTTTGATA GATCAAACTTGCCAT
       A   Q   E   N   L   A   E
1017 TTCATTCGGTATTTC GGTTGAAGTACGTAA CTAATATGCGTACGT GTGCATCGATCGATC AG | GG CGC AGC
                                                                       G   R   S
1087 CGC GGG GGG CAG ACG AGG AAG GAG CAG ATG GGG GAG GAA GGG TAC CGC GAG ATG GGG CGC
       R   G   G   Q   T   R   K   E   Q   M   G   E   E   G   Y   R   E   M   G   R
1147 AAG GGC GGC CTC AGC ACC GGC GAC GAG TCC GGC GGC GAG CGC GCC GCC CGC GAG GGC ATC
       K   G   G   L   S   T   G   D   E   S   G   G   E   R   A   A   R   E   G   I
1207 GAC ATC GAC GAG TCC AAG TAC AAG ACC AAG TCC TAG  ACTACACACACTTTT  GCATCCGTGAATG
       D   I   D   E   S   K   Y   K   T   K   S  Stop
1271 CCAGTGTGTCGTAGT CGTCTCAGTTGTAGT CATCGAGTCAGTCTT AGCTAGCTAGCTCTC TTATCAATAATAATG
1346 TAGGTCTTGACGGAT GCACGCATTTAGGCG CCCGTATGATTTGCT ATGTTATGTTTCATA TGGTGATGCGTGCTT
1421 AGCTTTAGCTAGCTT TGGTAGTTTGGTTTA GGTGCTGGTCAGTAG CTAGGGTATGTGTTT TAGTTTATTGCAGGA
1496 TCGGTCAGGCAGGGG CTGAATGCTAGCTAG GTGATGCTTTGGCTT TTCATGGCCAGCAGC TGCTAGCTTCTTGTT
1571 TCCAACTGATGCCTC TGCTGGGTGTGTCCT TTGTAACCGTGCTTG CTTTCAGTTAATCTA GCTAGCTGTTGTTCA
1646 TCCGAAAGTGCCAGT GTCTATGTCTGATCT ATCGCGGCTAATTGT TGCTGTTGCTAGTTG ATCTTTGATTCCTCA
1721 GAGATGCCATGTGCA GGATGGAATTATTAG CGCAAAACACAGGAG AAGGGACAACAATTT TAAGTTGCCTAATTT
1796 TTTTCTTAAAACACC TGCTGGGCCGTGCAC ACAAAAAAAAAATCA AGTACCTCGTCTAAA CTGAATGCTCAAGTT
1871 TGAAAGTGAAAACAG GACGTGCCGTGCCAA GAACTTCGAGATATC TCCGAGTTCT
```

Fig. 57

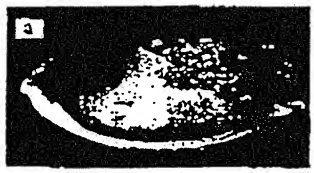


Fig. 58A

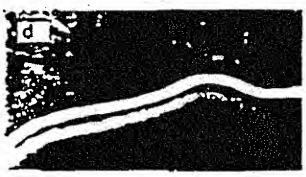


Fig. 58D

Fig. 58B

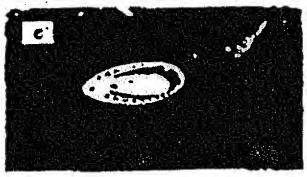


Fig. 58E

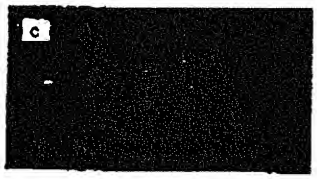


Fig. 58C

Fig. 58F

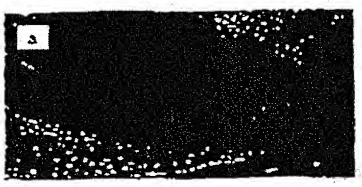


Fig. 59A

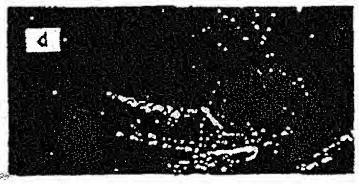


Fig. 59D

Fig. 59B

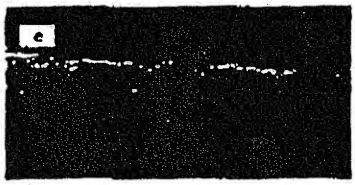


Fig. 59E

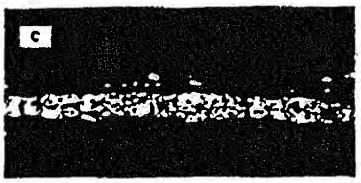


Fig. 59C

Fig. 59F

SEQUENCE LISTING

<110> Ventria Bioscience

<120> Expression System for Seed Proteins

<130> 50665-8026.WO00

<140> Not Yet Assigned
<141> Filed Herewith

<150> US 60/269,188
<151> 2001-02-14

<150> US 60/269,199
<151> 2001-02-14

<150> US 09/847,232
<151> 2001-05-02

<160> 33

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 393
<212> DNA
<213> Homo sapiens

<400> 1
aaagtcttcg agcggtgcga gctggcccgc acgctcaagc ggctcggcat ggacggctac 60
cggggcatca gcctcgccaa ctggatgtgc ctcgccaagt gggagtcggg ctacaacacc 120
cgcgcaacca actacaacgc cggcgaccgc tccaccgact acggcatctt ccagatcaac 180
tcccgctact ggtgcaacga cggcaagacg cccggggccg tcaacgcctg ccacctctcc 240
tgctcggccc tgctgcaaga caacatcgcc gacgccgtcg cgtgcgcgaa gcgcgtcgtc 300
cgcgacccgc agggcatccg ggcctgggtg gcctggcgca accgctgcca gaaccgggac 360
gtgcgccagt acgtccaggg ctgcggcgtc tga 393

<210> 2
<211> 130
<212> PRT
<213> Homo sapiens

<400> 2
Lys Val Phe Glu Arg Cys Glu Leu Ala Arg Thr Leu Lys Arg Leu Gly
1 5 10 15
Met Asp Gly Tyr Arg Gly Ile Ser Leu Ala Asn Trp Met Cys Leu Ala
20 25 30
Lys Trp Glu Ser Gly Tyr Asn Thr Arg Ala Thr Asn Tyr Asn Ala Gly
35 40 45
Asp Arg Ser Thr Asp Tyr Gly Ile Phe Gln Ile Asn Ser Arg Tyr Trp
50 55 60
Cys Asn Asp Gly Lys Thr Pro Gly Ala Val Asn Ala Cys His Leu Ser
65 70 75 80
Cys Ser Ala Leu Leu Gln Asp Asn Ile Ala Asp Ala Val Ala Cys Ala
85 90 95
Lys Arg Val Val Arg Asp Pro Gln Gly Ile Arg Ala Trp Val Ala Trp
100 105 110
Arg Asn Arg Cys Gln Asn Arg Asp Val Arg Gln Tyr Val Gln Gly Cys
115 120 125
Gly Val
130

<210> 3

<211> 2079
<212> DNA
<213> Homo sapiens

<400> 3
gggcggcggc ggcgctcggt gcagtggtgc gccgtgtccc agcccgaggc gaccaagtgc 60
ttccagtggc agcgcaacat gcggaaggtg cgcggcccgc cggtcagctg catcaagcgg 120
gactccccca tccaatgcat ccaggccatc gcggagaacc gcgccgacgc ggtcaccctg 180
gacggcgggt tcatctacga ggcggggctc gccccgtaca agctccgccc ggtggcggcg 240
gaggtgtacg gcaccgagcg ccagccgcgc acgcactact acgcggtggc cgtcgtcaag 300
aagggcgggt ccttccagct caacgagctg cagggcctga agtcgtgcca cacgggcctc 360
cggcggacgg cgggctggaa cgtgcccatc ggcaccctgc gccccttcct gaactggacc 420
ggcccgccgg agccgatcga ggccgccgtg gcccgcttct tcagcgcctc ctgcgtcccc 480
ggcgccgaca agggccagtt cccgaacctc tgccggctct gcgccgggac gggcgagaac 540
aagtgcgcct tctcctcgca ggagccgtac ttctcctact cgggcgcgtt caagtgcctc 600
cgcgacgggg ccggcgacgt ggcgttcatc cgcgagtcca ccgtgttcga ggacctctcc 660
gacgaggcgg agcgggacga gtacgagctg ctgtgccccg acaacacccg caagccggtg 720
gacaagttca aggactgcca cctggcgcgg gtgccctcgc acgcggtcgt cgcccgcagc 780
gtcaacggca aggaggacgc gatctggaac ctcctccgcc aggcccagga gaagttcggc 840
aaggacaagt cccccaagtt ccagctcttc gggagcccca gcggccagaa ggacctcctc 900
ttcaaggact ccgcgatcgg cttctcccgc gtccccccgc gcatcgactc cggcctgtac 960
ctcggctccg ggtacttcac cgcgatccag aacctccgga agagcgagga ggaggtggcg 1020
gcgcggcggg cccgcgtcgt gtggtgcgcc gtgggcgagc aggagctgcg gaagtgcaac 1080
cagtggagcg gcctgagcga ggggtcggtg acctgctcgt ccgccagcac caccgaggac 1140
tgcatcgcgc tcgtcctcaa gggggaggcc gacgcgatga gcctcgacgg ggggtacgtc 1200
tacaccgccg gcaagtgcgg cctggtcccg gtcctggcgg agaactacaa gtcgcagcag 1260
tccagcgacc ccgacccgaa ctgcgtggac cgccccgtcg agggctacct cgccgtggcc 1320
gtcgtgcgcc ggtccgacac ctccctgacg tggaacagcg tcaagggcaa gaagagctgc 1380
cacaccgccg tggaccgcac cgccggctgg aacatcccga tgggcctcct cttcaaccag 1440
accggctcct gcaagttcga cgagtacttc tcccagtcct gcgcccccgg ctcggacccc 1500
cgctccaacc tgtgcgccct ctgcatcggg gacgagcagg gcgagaacaa gtgcgtgccc 1560
aacagcaacg agcggtacta cggctacacg ggggccttcc gctgcctggc ggagaacgcc 1620
ggggacgtcg cgttcgtgaa ggacgtgacc gtgctgcaaa acacggacgg gaacaacaac 1680
gaggcgtggg cgaaggacct caagctcgcc gacttcgccc tgctgtgcct cgacggcaag 1740
cgcaagcccg tcaccgaggc gcggtcctgc cacctggcga tggcccccaa ccacgccgtc 1800
gtctcccgca tggacaaggt cgagcgcctc aagcaggtgc tcctgcacca gcaggccaag 1860
ttcggccgga acggcagcga ctgcccggac aagttctgcc tgttccagtc ggagaccaag 1920
aacctcctct tcaacgacaa caccgagtgc ctggcgcgcc tccacggcaa gaccacctac 1980
gagaagtacc tcggcccgca gtacgtcgcc ggcatcacca acctcaagaa gtgctccacc 2040
tccccctcc tggaggcgtg cgagttcctc cgcaagtga 2079

<210> 4
<211> 690
<212> PRT
<213> Homo sapiens

<400> 4
Gly Arg Arg Arg Arg Ser Val Gln Trp Cys Ala Val Ser Gln Pro Glu
1 5 10 15
Ala Thr Lys Cys Phe Gln Trp Gln Arg Asn Met Arg Lys Val Arg Gly
20 25 30
Pro Pro Val Ser Cys Ile Lys Arg Asp Ser Pro Ile Gln Cys Ile Gln
35 40 45
Ala Ile Ala Glu Asn Arg Ala Asp Ala Val Thr Leu Asp Gly Gly Phe
50 55 60
Ile Tyr Glu Ala Gly Leu Ala Pro Tyr Lys Leu Arg Pro Val Ala Ala
65 70 75 80
Glu Val Tyr Gly Thr Glu Arg Gln Pro Arg Thr His Tyr Tyr Ala Val
85 90 95
Ala Val Val Lys Lys Gly Gly Ser Phe Gln Leu Asn Glu Leu Gln Gly
100 105 110
Leu Lys Ser Cys His Thr Gly Leu Arg Arg Thr Ala Gly Trp Asn Val
115 120 125
Pro Ile Gly Thr Leu Arg Pro Phe Leu Asn Trp Thr Gly Pro Pro Glu
130 135 140

```
Pro Ile Glu Ala Ala Val Ala Arg Phe Phe Ser Ala Ser Cys Val Pro
145                 150                 155                 160
Gly Ala Asp Lys Gly Gln Phe Pro Asn Leu Cys Arg Leu Cys Ala Gly
                165                 170                 175
Thr Gly Glu Asn Lys Cys Ala Phe Ser Ser Gln Glu Pro Tyr Phe Ser
            180                 185                 190
Tyr Ser Gly Ala Phe Lys Cys Leu Arg Asp Gly Ala Gly Asp Val Ala
        195                 200                 205
Phe Ile Arg Glu Ser Thr Val Phe Glu Asp Leu Ser Asp Glu Ala Glu
    210                 215                 220
Arg Asp Glu Tyr Glu Leu Leu Cys Pro Asp Asn Thr Arg Lys Pro Val
225                 230                 235                 240
Asp Lys Phe Lys Asp Cys His Leu Ala Arg Val Pro Ser His Ala Val
                245                 250                 255
Val Ala Arg Ser Val Asn Gly Lys Glu Asp Ala Ile Trp Asn Leu Leu
            260                 265                 270
Arg Gln Ala Gln Glu Lys Phe Gly Lys Asp Lys Ser Pro Lys Phe Gln
        275                 280                 285
Leu Phe Gly Ser Pro Ser Gly Gln Lys Asp Leu Leu Phe Lys Asp Ser
    290                 295                 300
Ala Ile Gly Phe Ser Arg Val Pro Pro Arg Ile Asp Ser Gly Leu Gly
305                 310                 315                 320
Ser Gly Tyr Phe Thr Ala Ile Gln Asn Leu Arg Lys Ser Glu Glu Glu
                325                 330                 335
Val Ala Ala Arg Arg Ala Arg Val Val Trp Cys Ala Val Gly Glu Gln
            340                 345                 350
Glu Leu Arg Lys Cys Asn Gln Trp Ser Gly Leu Ser Glu Gly Ser Val
        355                 360                 365
Thr Cys Ser Ser Ala Ser Thr Thr Glu Asp Cys Ile Ala Leu Val Leu
    370                 375                 380
Lys Gly Glu Ala Asp Ala Met Ser Leu Asp Gly Gly Tyr Val Tyr Thr
385                 390                 395                 400
Ala Gly Lys Cys Gly Leu Val Pro Val Leu Ala Glu Asn Tyr Lys Ser
                405                 410                 415
Gln Gln Ser Ser Asp Pro Asp Pro Asn Cys Val Asp Arg Pro Val Glu
            420                 425                 430
Gly Tyr Leu Ala Val Ala Val Val Arg Arg Ser Asp Thr Ser Leu Thr
        435                 440                 445
Trp Asn Ser Val Lys Gly Lys Lys Ser Cys His Thr Ala Val Asp Arg
    450                 455                 460
Thr Ala Gly Trp Asn Ile Pro Met Gly Leu Leu Phe Asn Gln Thr Gly
465                 470                 475                 480
Ser Cys Lys Phe Asp Glu Tyr Phe Ser Gln Ser Cys Ala Pro Gly Ser
                485                 490                 495
Asp Pro Arg Ser Asn Leu Cys Ala Leu Cys Ile Gly Asp Glu Gln Gly
            500                 505                 510
Glu Asn Lys Cys Val Pro Asn Ser Asn Glu Arg Tyr Tyr Gly Tyr Thr
        515                 520                 525
Gly Ala Phe Arg Cys Leu Ala Glu Asn Ala Gly Asp Val Ala Phe Val
    530                 535                 540
Lys Asp Val Thr Val Leu Gln Asn Thr Asp Gly Asn Asn Asn Glu Ala
545                 550                 555                 560
Trp Ala Lys Asp Leu Lys Leu Ala Asp Phe Ala Leu Leu Cys Leu Asp
                565                 570                 575
Gly Lys Arg Lys Pro Val Thr Glu Ala Arg Ser Cys His Leu Ala Met
            580                 585                 590
Ala Pro Asn His Ala Val Val Ser Arg Met Asp Lys Val Glu Arg Leu
        595                 600                 605
Lys Gln Val Leu Leu His Gln Gln Ala Lys Phe Gly Arg Asn Gly Ser
    610                 615                 620
Asp Cys Pro Asp Lys Phe Cys Leu Phe Gln Ser Glu Thr Lys Asn Leu
625                 630                 635                 640
Leu Phe Asn Asp Asn Thr Glu Cys Leu Ala Arg Leu His Gly Lys Thr
                645                 650                 655
Thr Tyr Glu Lys Tyr Leu Gly Pro Gln Tyr Val Ala Gly Ile Thr Asn
```

660 665 670
Leu Lys Lys Cys Ser Thr Ser Pro Leu Leu Glu Ala Cys Glu Phe Leu
675 680 685
Arg Lys
690

<210> 5
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 5
atcgaagctt catgagtaat gtgtgagcat tatgggacca cg 42

<210> 6
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 6
ctagtctaga ctcgagccat ggggccggct agggagccat cgcacaagag gaa 53

<210> 7
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> codon optimized lactoferricin coding sequence based on Homo sapiens sequence

<400> 7
accaagtgct tccagtggca gcgcaacatg cggaaggtgc gcggcccgcc ggtcagctgc 60
atcaagcggg ac 72

<210> 8
<211> 162
<212> DNA
<213> Artificial Sequence

<220>
<223> codon optimized EGF coding sequence based on Homo sapiens sequence

<400> 8
aactccgact cggagtgccc cctctcccac gacggttact gcctccacga cggggtctgc 60
atgtacatcg aggccctcga caagtacgcc tgcaactgcg tcgtgggcta catcggcgag 120
cggtgccagt accgcgacct caagtggtgg gagctgcgct ga 162

<210> 9
<211> 213
<212> DNA
<213> Artificial Sequence

<220>
<223> codon optimized IGF-1 coding sequence based on Homo sapiens sequence

<400> 9
ggcccggaga ccctctgcgg cgccgagctc gtggacgccc tccagttcgt gtgcggcgac 60

cgcggcttct acttcaacaa gccgaccggc tacggcagca gcagccgccg cgccccgcag 120
accggcatcg tggacgagtg ctgcttccgc agctgcgacc tccgccgcct ggagatgtac 180
tgcgccccgc tcaagcccgc caagagcgcc tga 213

<210> 10
<211> 1095
<212> DNA
<213> Artificial Sequence

<220>
<223> codon optimized lactohedrin coding sequence based on Homo sapiens sequence

<400> 10
ctggacatct gctcgaagaa cccgtgccac aacggcgggc tctgcgagga gatcagccag 60
gaggtgcggg gcgacgtgtt cccctcgtac acctgcacct gcctgaaggg ctacgccggg 120
aaccactgcg agacgaagtg cgtggagccc ctggggatgg agaacggcaa catcgccaac 180
tcccagatcg ccgcctcctc cgtgcgggtg accttcctcg gcctccagca ctgggtcccg 240
gagctggccc ggctcaaccg ggcgggcatg gtgaacgcgt ggacccccctc gtccaacgac 300
gacaacccgt ggatccaagt gaacctgctc cgccgcatgt gggtcaccgg cgtggtcacc 360
caaggcgcca gccgcctggc cagccacgag tacctcaagg ccttcaaggt cgcctacagc 420
ctcaacggcc acgagttcga cttcatccac gacgtcaaca agaagcacaa ggagttcgtg 480
ggcaactgga acaagaacgc ggtccacgtg aacctcttcg agacccccgt cgaggcccag 540
tacgtccgcc tctaccccac gagctgccac accgcctgca cgctccgctt cgagctgctg 600
gggtgcgagc tgaacgggtg cgcgaacccg ctggggctca agaacaacag catccccgac 660
aagcagatca cggcctcgtc gtcgtacaag acctggggcc tgcacctctt ctcgtggaac 720
ccgagctacg cccggctgga caagcagggc aacttcaacg cctgggtcgc cgggagctac 780
gggaacgacc agtggctcca ggtggacctc ggcagctcca aggaggtcac cggcatcatc 840
acgcaggggg cccgcaactt cggctccgtg cagttcgtgg cctcctacaa ggtggcctac 900
tcgaacgaca gcgccaactg gaccgagtac caggacccgc gcaccgggtc cagcaagatc 960
ttccccggca actgggacaa ccacagccac aagaagaacc tgttcgagac ccccatcctc 1020
gcccggtacg tccgcatcct ccccgtcgct tggcacaacc ggatcgcgct ccggctggag 1080
ctcctcggct gctga 1095

<210> 11
<211> 489
<212> DNA
<213> Artificial Sequence

<220>
<223> codon optimized kappa-casein coding sequence based on Homo sapiens sequence

<400> 11
gaggtccaaa accagaagca gcccgcctgc cacgagaacg acgagcgccc cttctaccag 60
aagaccgcac cctacgtccc gatgtactac gtcccgaaca gctaccccta ctacggtacg 120
aacctgtacc agcgccgccc ggccatcgct atcaacaacc cctacgtccc ccggacctac 180
tacgcgaacc cggccgtggt gcggccccac gcgcagatcc cgcagcggca gtacctgcca 240
aacagccacc cccccaccgt ggtgcggcgg cccaacctcc acccgagctt catcgctatc 300
ccccccaaga agatccagga caagatcatc atcccgacca tcaacaccat cgccaccgtg 360
gagccgacgc cagcccccgc gaccgagccc acggtggaca gcgtcgtgac cccagaggcg 420
ttctccgaat cgatcatcac ctccaccccc gagaccacca cggtggccgt cacgccgccg 480
acggcatga 489

<210> 12
<211> 1233
<212> DNA
<213> Artificial Sequence

<220>
<223> codon optimized haptocorrin coding sequence based on Homo sapiens sequence

<400> 12
gagatctgcg aggtctccga ggagaactac atccgcctca agcccctcct gaacaccatg 60

atccagagca actacaaccg gggcacgtcg gccgtgaacg tcgtgctctc cctgaagctc 120
gtgggcatcc agatccagac cctcatgcag aagatgatcc agcagatcaa gtacaacgtg 180
aagagccgcc tctcggacgt gtccagcggc gagctggcgc tcatcatcct cgcgctcggc 240
gtgtgccgga acgcggagga gaacctcatc tacgactacc acctcacgga caagctggag 300
aacaagttcc aggccgagat cgagaacatg gaggcccaca acggcacccc gctgaccaac 360
tactaccagc tcagcctgga cgtcctcgcg ctctgcctgt tcaacgggaa ctactccacc 420
gccgaggtgg tcaaccactt caccccgag aacaagaact actacttcgg ctcgcagttc 480
tccgtggaca ccggggccat ggccgtcctg gccctcacct gcgtgaagaa gtccctcatc 540
aacggccaga tcaaggccga cgagggctcc ctgaagaaca tctcgatcta caccaagagc 600
ctcgtggaga agatcctcag cgagaagaag gagaacgggc tgatcggcaa caccttctcg 660
accggcgagg cgatgcaggc cctgttcgtg agcagcgact actacaacga gaacgactgg 720
aactgccagc agaccctcaa cacggtcctg accgagatca gccagggcgc gttcagcaac 780
cccaacgccg ccgcccaggt cctgccggcc ctgatgggca agaccttcct cgacatcaac 840
aaggacagct cctgcgtgtc cgcgagcggc aacttcaaca tctccgccga cgagccgatc 900
acggtgacgc cgcccgacag ccagtcgtac atctccgtga actacagcgt gcggatcaac 960
gagacctact tcacgaacgt gacggtcctc aacggctcgg tcttcctgag cgtgatggag 1020
aaggcgcaga agatgaacga cacgatcttc ggcttcacga tggaggagcg cagctggggc 1080
ccctacatca cctgcatcca gggcctctgc gccaacaaca acgaccgcac ctactgggag 1140
ctgctgagcg gcggcgagcc gctgagccag ggggccggca gctacgtggt ccgcaacggc 1200
gagaacctgg aggtccggtg gagcaagtac tga 1233

<210> 13
<211> 2061
<212> DNA
<213> Artificial Sequence

<220>
<223> codon optimized lactoperoxidase coding sequence based on Homo sapiens sequence

<400> 13
caaacgaccc ggacgtcggc gatctccgac acggtctcgc aggccaaggt gcaagtcaac 60
aaggcattcc tggattcgcg cacgcggctg aagaccgcga tgtcgtccga gaccccgacg 120
agccggcagc tgagcgagta cctcaagcac gcgaaggggc ggacgcgcac cgccatccgc 180
aatggccaag tgtgggagga atccctgaag cggctgcggc agaaggcgtc gctcaccaac 240
gtgaccgacc cgtccctcga cctgaccagc ctctccctgg aggtcggctg cggcgccccg 300
gcgcccgtcg tgcgctgcga cccctgctcg ccataccgca cgatcacggg cgactgcaac 360
aaccggcgga agccggcact gggggctgcg aaccgcgccc tcgcgcgctg gctccccgcc 420
gagtacgagg acggcctcag cctccccttc ggttggaccc ccggcaagac gcgcaacggc 480
ttcccgctcc cgctcgctcg cgaggtcagc aacaagatcg tcggttacct gaacgaggag 540
ggggtcctcg accaaaaccg ctccctcctc ttcatgcagt gggggcagat cgtggaccac 600
gacctggact tcgccccgga cacggagctg ggctccagcg agtacagcaa gacccagtgc 660
gacgaatact gcatccaggg cgacaactgc ttcccgatca tgttcccccc gaacgacccg 720
aaggcgggca cccagggcaa gtgcatgccg ttcttccggg caggcttcgt ctgcccgacc 780
cccccgtaca agtccctcgc gcgcgagcag atcaacgcgc tcacgtcctt cctcgacgcc 840
agcttcgtct acagcagcga gccgtccctc gccagccgcc tccgcaacct cagcagcccc 900
ctcggcctca tggcggtcaa ccaggaggtg tcggaccacg gcctcccata cctgccgtac 960
gacagcaaga agccgtcccc ctgcgagttc atcaacacca ccgcgcgcgt cccgtgcttc 1020
ctcgccggcg attcgcgggc gagcgagcac atcctcctcg ccacgagcca caccctgttc 1080
ctccgcgagc acaaccgcct cgcccgggag ctgaagcgcc tcaacccgca gtgggacggc 1140
gagaagctct accaggaggc ccggaagatc ctcggcgctt tcgtccagat catcaccttc 1200
cgggactacc tcccccatcct gctcggtgac cacatgcaga agtggatccc cccctaccaa 1260
ggctactccg agagcgtgga cccgcgcatc tccaacgtct tcacgttcgc gttccgcttc 1320
gggcacctgg aggtgccgtc gtcgatgttc cgcctcgacg agaactacca gccctggggc 1380
ccagagccgg agctgccgct ccacaccctg ttcttcaaca cctggcggat ggtcaaggac 1440
ggcggcatcg acccgctcgt gcgcgggctc ctggctaaga agtcgaagct catgaagcag 1500
aacaagatga tgaccggcga gctgcgcaac aagctgttcc agcccaccca ccgcatccac 1560
gggttcgacc tggctgcaat caacacccag cggtgccgcg accacggcca gcccggctac 1620
aactcgtggc gcgcgttctg cgacctctcc cagccacaga cgctggagga gctcaacacc 1680
gtgctcaaga gcaagatgct cgccaagaag ctgctcgggc tctacggcac gcccgacaac 1740
atcgacatct ggatcggggc catcgcggag ccgctcgtgg agcgcgggcg cgtcggcccg 1800
ctgctcgcgt gcctcctggg caagcaattc caacagatcc gcgacgggga ccggttctgg 1860
tgggagaacc ccggcgtgtt caccaacgag cagaaggatt cgctccaaaa gatgagcttc 1920
tcccgcctgg tgtgcgacaa cacccgcatc accaaggtcc cgcgcgaccc attctgggcc 1980
aactcctacc cgtacgactt cgtggactgc tccgccatcg acaagctcga cctgtccccc 2040

tgggcatcgg tgaagaactg a 2061

<210> 14
<211> 1185
<212> DNA
<213> Artificial Sequence

<220>
<223> codon optimized alpha-1-antitrypsin coding sequence based on Homo sapiens sequence

<400> 14
gaggacccgc agggcgacgc cgcccagaag accgacacca gccaccacga ccaggaccac 60
ccgacgttca acaagatcac cccgaatttg gccgaattcg ccttcagcct gtaccgccag 120
ctcgcgcacc agtccaactc caccaacatc ttcttcagcc cggtgagcat cgccaccgcc 180
ttcgccatgc tgtccctggg taccaaggcg gacacccacg acgagatcct cgaagggctg 240
aacttcaacc tgacggagat cccggaggcg cagatccacg agggcttcca ggagctgctc 300
aggacgctca accagccgga ctcccagctc cagctcacca ccggcaacgg gctcttcctg 360
tccgagggcc tcaagctcgt cgataagttc ctggaggacg tgaagaagct ctaccactcc 420
gaggcgttca ccgtcaactt cggggacacc gaggaggcca agaagcagat caacgactac 480
gtcgagaagg ggacccaggg caagatcgtg gacctggtca aggaattgga cagggacacc 540
gtcttcgcgc tcgtcaacta catcttcttc aagggcaagt gggagcgccc gttcgaggtg 600
aaggacaccg aggaggagga cttccacgtc gaccaggtca ccaccgtcaa ggtcccgatg 660
atgaagaggc tcggcatgtt caacatccag cactgcaaga agctctccag ctgggtgctc 720
ctcatgaagt acctggggaa cgccaccgcc atcttcttcc tgccggacga gggcaagctc 780
cagcacctgg agaacgagct gacgcacgac atcatcacga agttcctgga gaacgaggac 840
aggcgctccg ctagcctcca cctcccgaag ctgagcatca ccggcacgta cgacctgaag 900
agcgtgctgg gccagctggg catcacgaag gtcttcagca acggcgcgga cctctccggc 960
gtgacggagg aggccccctt gaagctctcc aaggccgtgc acaaggcggt gctcacgatc 1020
gacgagaagg ggacggaagc tgccggggcc atgttcctgg aggccatccc cgtgtccatc 1080
ccgcccgagg tcaagttcaa caagcccttc gtcttcctga tgatcgagca gaacacgaag 1140
agccccctct tcatggggaa ggtcgtcaac cccacgcaga agtga 1185

<210> 15
<211> 786
<212> DNA
<213> Artificial Sequence

<220>
<223> Rice Gt1 promoter and Gt1 leader coding sequence

<400> 15
catgagtaat gtgtgagcat tatgggacca cgaaataaaa agaacatttt gatgagtcgt 60
gtatcctcga tgagcctcaa aagttctctc accccggata agaaaccctt aagcaatgtg 120
caaagtttgc attctccact gacataatgc aaaataagat atcatcgatg acatagcaac 180
tcatgcatca tatcatgcct ctctcaacct attcattcct actcatctac ataagtatct 240
tcagctaaat gttagaacat aaacccataa gtcacgtttg atgagtatta ggcgtgacac 300
atgacaaatc acagactcaa gcaagataaa gcaaaatgat gtgtacataa aactccagag 360
ctatatgtca tattgcaaaa agaggagagc ttataagaca aggcatgact cacaaaaatt 420
cacttgcctt tcgtgtcaaa aagaggaggg ctttacatta tccatgtcat attgcaaaag 480
aaagagagaa agaacaacac aatgctgcgt caattataca tatctgtatg tccatcatta 540
ttcatccacc tttcgtgtac cacacttcat atatcataag agtcacttca cgtctggaca 600
ttaacaaact ctatcttaac atttagatgc aagagccttt atctcactat aaatgcacga 660
tgatttctca ttgtttctca caaaaagcgg ccgcttcatt agtcctacaa caacatggca 720
tccataaatc gccccatagt tttcttcaca gtttgcttgt tcctcttgtg cgatggctcc 780
ctagcc 786

<210> 16
<211> 1055
<212> DNA
<213> Artificial Sequence

<220>
<223> Rice Glb promoter and Gt1 leader coding sequence

<400> 16
ctgcagggag gagaggggag agatggtgag agaggaggaa gaagaggagg ggtgacaatg 60
atatgtgggg catgtgggca cccaattttt taattcattc ttttgttgaa actgacatgt 120
gggtcccatg agatttatta tttttcggat cgaatcgcca cgtaagcgct acgtcaatgc 180
tacgtcagat gaagaccgag tcaaattagc cacgtaagcg ccacgtcagc caaaaccacc 240
atccaaaccg ccgagggacc tcatctgcac tggttttgat agttgaggga cccgttgtat 300
ctggtttttc gattgaagga cgaaaatcaa atttgttgac aagttaaggg accttaaatg 360
aacttattcc atttcaaaat attctgtgag ccatatatac cgtgggcttc caatcctcct 420
caaattaaag ggccttttta aaatagataa ttgccttctt tcagtcaccc ataaaagtac 480
aaaactacta ccaacaagca acatgcgcag ttacacacat tttctgcaca tttccgccac 540
gtcacaaaga gctaagagtt atccctagga caatctcatt agtgtagata catccattaa 600
tcttttatca gaggcaaacg taaagccgct ctttatgaca aaaataggtg acacaaaagt 660
gttatctgcc acatacataa cttcagaaat tacccaacac caagagaaaa ataaaaaaaa 720
atctttttgc aagctccaaa tcttggaaac ctttttcact ctttgcagca ttgtactctt 780
gctctttttc caaccgatcc atgtcaccct caagcttcta cttgatctac acgaagctca 840
ccgtgcacac aaccatggcc acaaaaaccc tataaaaccc catccgatcg ccatcatctc 900
atcatcagtt cattaccaac aaacaaaaga ggaaaaaaaa catatacact tctagtgatt 960
gtctgattga tcatcaatct agaggcggcc gcatggctag caaggtcgtc ttcttcgcgg 1020
cggcgctcat ggcggccatg gtggccatct ccggc 1055

<210> 17
<211> 976
<212> DNA
<213> Artificial Sequence

<220>
<223> Bx7 promoter

<400> 17
ctgcaggcca gggaaagaca atggacatgc aaagaggtag gggcagggaa gaaacacttg 60
gagatcatag aagaacataa gaggttaaac ataggagggc ataatggaca attaaatcta 120
cattaattga actcatttgg gaagtaaaca aaatccatat tctggtgtaa atcaaactat 180
ttgacgcgga tttactaaga tcctatgtta attttagaca tgactggcca aaggtttcag 240
ttagttcatt tgtcacggaa aggtgttttc ataagtccaa aactctacca actttttgc 300
acgtcatagc atagatagat gttgtgagtc attggataga tattgtgagt cagcatggat 360
ttgtgttgcc tggaaatcca actaaatgac aagcaacaaa acctgaaatg ggctttagga 420
gagatggttt atcaatttac atgttccatg caggctacct tccactactc gacatggtta 480
gaagttttga gtgccgcata tttgcggaag caatggcact actcgacatg gttagaagtt 540
ttgagtgccg catatttgcg gaagcaatgg ctaacagata catattctgc caaaccccaa 600
gaaggataat cactcctctt agataaaaag aacagaccaa tgtacaaaca tccacacttc 660
tgcaaacaat acaccagaac taggattaag cccattacgt ggctttagca gaccgtccaa 720
aaatctgttt tgcaagcacc aattgctcct tacttatcca gcttcttttg tgttggcaaa 780
ctgccctttt ccaaccgatt ttgtttcttc tcacgctttc ttcataggct aaactaacct 840
cggcgtgcac acaaccatgt cctgaacctt cacctcgtcc ctataaaagc ccatccaacc 900
ttacaatctc atcatcaccc acaacaccga gcaccccaat ctacagatca attcactgac 960
agttcactga tctaga 976

<210> 18
<211> 1009
<212> DNA
<213> Artificial Sequence

<220>
<223> Glub-2 promoter

<400> 18
ctgcagtaat ggatacctag tagcaagcta gcttaaacaa atctaaattc caatctgttc 60
gtaaacgttt tctcgatcgc aattttgatc aaaactattg aaaacctcaa ttaaaccatt 120
caaaatttttt aatataccca acaagagcgt ccaaaccaaa tatgtaaata tggatgtcat 180
gataattgac ttatgacaat gtgattattt catcaagtct ttaaatcatt aattctagtt 240
gaaggtttat gttttcttat gctaaagggt tatgtttata taagaatatt aaagagcaaa 300
ttgcaataga tcaacacaac aaatttgaat gtttccagat gtgtaaaaat atccaaatta 360
attgttttaa aatagtttta agaaggatct gatatgcaag tttgatagtt agtaaactgc 420
aaaagggctt attacatgga aaattcctta ttgaatatgt ttcattgact ggtttatttt 480
acatgacaac aaagttacta gtatgtcaat aaaaaaatac aaggttactt gtcaattgta 540

ttgtgccaag taaagatgac aacaaacata caaatttatt tgttctttta tagaaacacc 600
taacttatca aggatagttg gccacgcaaa aatgacaaca tactttacaa ttgtatcatc 660
ataaagatct tatcaagtat aagaacttta tggtgacata aaaaataatc acaagggcaa 720
gacacatact aaaagtatgg acagaaattt cttaacaaac tccatttgtt ttgtatccaa 780
aagcataaga aatgagtcat ggctgagtca tgatatgtag ttcaatcttg caaaattgcc 840
tttttgttaa gtattgtttt aacactacaa gtcacatatt gtctatactt gcaacaaaca 900
ctattaccgt gtatcccaag tggccttttc attgctatat aaactagctt gatcggtctt 960
tcaactcaca tcaattagct taagtttcca ttagcaactg ctaatagct 1009

<210> 19
<211> 839
<212> DNA
<213> Artificial Sequence

<220>
<223> Gt3 promoter

<400> 19
ctgcagtgta agtgtagctt cttatagctt agtgctttac tatcttcaca agcacatgct 60
atagtattgt tccaagatga aagaataatt catccttgct accaacttgc atgatattat 120
atttgtgaat atcctatctc ttggcttata atgaaatgtg ctgctgggtt attctgacca 180
tggtatttga gagcctttgt atagctgaaa ccaacgtata tcgagcatgg aacagagaac 240
aaaatgcaag gatttttta ttctggttca tgccctggat gggttaatat cgtgatcatc 300
aaaaaagata tgcataaaat taaagtaata aatttgctca taagaaacca aaaccaaaag 360
cacatatgtc ctaaacaaac tgcattttgt ttgtcatgta gcaatacaag agataatata 420
tgacgtggtt atgacttatt cacttttttgt gactccaaaa tgtagtaggt ctaactgatt 480
gtttaaagtg atgtcttact gtagaagttt catcccaaaa gcaatcacta aagcaacaca 540
cacgtatagt ccaccttcac gtaattcttt gtggaagata acaagaaggc tcactgaaaa 600
ataaaagcaa agaaaaggat atcaaacaga ccattgtgca tcccattgat ccttgtatgt 660
ctatttatct atcctccttt tgtgtacctt acttctatct agtgagtcac ttcatatgtg 720
gacattaaca aactctatct taacatctag tcgatcacta ctttacttca ctataaaagg 780
accaacatat atcatccatt tctcacaaaa gcattgagtt cagtcccaca aaatctaga 839

<210> 20
<211> 1302
<212> DNA
<213> Artificial Sequence

<220>
<223> Glub-1 promoter

<400> 20
ctgcagagat atggattttc taagattaat tgattctctg tctaaagaaa aaaagtatta 60
ttgaattaaa tggaaaaaga aaaaggaaaa aggggatggc ttctgctttt tgggctgaag 120
gcggcgtgtg gccagcgtgc tgcgtgcgga cagcgagcga acacacgacg gagcagctac 180
gacgaacggg ggaccgagtg gaccggacga ggatgtggcc taggacgagt gcacaaggct 240
agtggactcg gtccccgcgc ggtatcccga gtggtccact gtctgcaaac acgattcaca 300
tagagcgggc agacgcggga gccgtcctag gtgcaccgga agcaaatccg tcgcctgggt 360
ggatttgagt gacacggccc acgtgtagcc tcacagctct ccgtggtcag atgtgtaaaa 420
ttatcataat atgtgttttt caaatagtta aataatatat ataggcaagt tatatgggtc 480
aataagcagt aaaaaggctt atgacatggt aaaattactt acaccaatat gccttactgt 540
ctgatatatt ttacatgaca acaaagttac aagtacgtca tttaaaaata caagttactt 600
atcaattgta gtgtatcaag taaatgacaa caaacctaca aatttgctat tttgaaggaa 660
cacttaaaaa aatcaatagg caagttatat agtcaataaa ctgcaagaag gcttatgaca 720
tggaaaaatt acatacacca atatgcttta ttgtccggta tattttacaa gacaacaaag 780
ttataagtat gtcatttaaa aatacaagtt acttatcaat tgtcaagtaa atgaaaacaa 840
acctacaaat ttgttatttt gaaggaacac ctaaattatc aaatatagct tgctacgcaa 900
aatgacaaca tgcttacaag ttattatcat cttaaagtta gactcatctt ctcaagcata 960
agagctttat ggtgcaaaaa caaatataat gacaaggcaa agatacatac atattaagag 1020
tatggacaga catttcttta acaaactcca tttgtattac tccaaaagca ccagaagttt 1080
gtcatggctg agtcatgaaa tgtatagttc aatcttgcaa agttgccttt ccttttgtac 1140
tgtgttttaa cactacaagc catatattgt ctgtacgtgc aacaaactat atcaccatgt 1200
atcccaagat gctttttttat tgctatataa actagcttgg tctgtctttg aactcacatc 1260
aattagctta agtttccata agcaagtaca aatagctcta ga 1302

<210> 21
<211> 675
<212> DNA
<213> Artificial Sequence

<220>
<223> Rice proalmin promoter

<400> 21
ctgcagcatc ggcttaggtg tagcaacacg actttattat tattattatt attattatta 60
ttatttttaca aaaatataaa atagatcagt ccctcaccac aagtagagca agttggtgag 120
ttattgtaaa gttctacaaa gctaatttaa aagttattgc attaacttat ttcatattac 180
aaacaagagt gtcaatggaa caatgaaaac catatgacat actataattt tgtttttatt 240
attgaaatta tataattcaa agagaataaa tccacatagc cgtaaagttc tacatgtggt 300
gcattaccaa aatatatata gcttacaaaa catgacaagc ttagtttgaa aaattgcaat 360
ccttatcaca ttgacacata aagtgagtga tgagtcataa tattattttt cttgctaccc 420
atcatgtata tatgatagcc acaaagttac tttgatgatg atatcaaaga acatttttag 480
gtgcacctaa cagaatatcc aaataatatg actcacttag atcataatag agcatcaagt 540
aaaactaaca ctctaaagca accgatggga aagcatctat aaatagacaa gcacaatgaa 600
aatcctcatc atccttcacc acaattcaaa tattatagtt gaagcatagt agtagaatcc 660
aacaacaatc tagag 675

<210> 22
<211> 1098
<212> DNA
<213> Artificial Sequence

<220>
<223> Rice cysteine peptidase promoter

<400> 22
ccaggcttca tcctaaccat tacaggcaag atgttgtatg aagaagggcg aacatgcaga 60
ttgttaaact gacacgtgat ggacaagaat gaccgattgg tgaccggtct gacaatggtc 120
atgtcgtcag cagacagcca tctcccacgt cgcgcctgct tccggtgaaa gtggaggtag 180
gtatgggccg tcccgtcaga aggtgattcg gatggcagcg atacaaatct ccgtccatta 240
atgaagagaa gtcaagttga aagaaaggga gggagagatg gtgcatgtgg gatccccttg 300
ggatataaaa ggaggacctt gcccacttag aaaggagagg agaaagcaat cccagaagaa 360
tcgggggctg actggcactt tgtagcttct tcatacgcga atccaccaaa acacaggagt 420
agggtattac gcttctcagc ggcccgaacc tgtatacatc gcccgtgtct tgtgtgtttc 480
cgctcttgcg aaccttccac agattgggag cttagaacct cacccagggc ccccggccga 540
actggcaaag gggggcctgc gcggtctccc ggtgaggagc cccacgctcc gtcagttcta 600
aattacccga tgagaaaggg agggggggggg gggaaatctg ccttgtttat ttacgatcca 660
acggatttgg tcgacaccga tgaggtgtct taccagttac cacgagctag attatagtac 720
taattacttg aggattcggt tcctaatttt ttacccgatc gacttcgcca tggaaaattt 780
tttattcggg ggagaatatc caccctgttt cgctcctaat taagatagga attgttacga 840
ttagcaacct aattcagatc agaattgtta gttagcggcg ttggatccct cacctcatcc 900
catcccaatt cccaaaccca aactcctctt ccagtcgccg acccaaacac gcatccgccg 960
cctataaatc ccacccgcat cgagcctatc aagcccaaaa aaccacaaac caaacgaaga 1020
aggaaaaaaa aaggaggaaa agaaaagagg aggaaagcga agaggttgga gagagacgct 1080
cgtctccacg tcgccgcc 1098

<210> 23
<211> 432
<212> DNA
<213> Artificial Sequence

<220>
<223> Barley D-Hordian promoter

<400> 23
cttcgagtgc ccgccgattt gccagcaatg gctaacagac acatattctg ccaaaacccc 60
agaacaataa tcacttctcg tagatgaaga gaacagacca agatacaaac gtccacgctt 120
cagcaaacag taccccagaa ctaggattaa gccgattacg cggctttagc agaccgtcca 180
aaaaaactgt tttgcaaagc tccaattcct ccttgcttat ccaatttctt ttgtgttggc 240
aaactgcact tgtccaaccg attttgttct tcccgtgttt cttcttaggc taactaacac 300

agccgtgcac atagccatgg tccggaatct tcacctcgtc cctataaaag cccagccaat 360
ctccacaatc tcatcatcac cgagaacacc gagaaccaca aaactagaga tcaattcatt 420
gacagtccac cg 432

<210> 24
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> bx7 signal peptide sequence

<400> 24
atggctaagc gcctggtcct ctttgcggca gtagtcgtcg ccctcgtggc tctcaccgcc 60

<210> 25
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> Glub-2 signal peptide sequence

<400> 25
atggcaacta ccattttctc tcgtttttct atatactttt gtgctatgct attatgccag 60
ggttctatgg cc 72

<210> 26
<211> 85
<212> DNA
<213> Artificial Sequence

<220>
<223> Gt3 signal peptide sequence

<400> 26
atgtggacat taacaaactc tatcttaaca tctagtcgat cactacttta cttcactata 60
aaaggaccaa catatatcat ccatt 85

<210> 27
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> Glub-2 signal peptide sequence

<400> 27
atggcgagtt ccgttttctc tcggttttct atatactttt gtgttcttct attatgccat 60
ggttctatgg cc 72

<210> 28
<211> 69
<212> DNA
<213> Artificial Sequence

<220>
<223> proalmin signal peptide sequence

<400> 28
atgaagatca ttttcgtatt tgctctcctt gctattgttg catgcaacgc ttctgcacgg 60
tttgatgct 69

<210> 29
<211> 63

<212> DNA
<213> Artificial Sequence

<220>
<223> Rice cysteine peptidase signal peptide sequence

<400> 29
atggccgccc gcgccgccgc cgccgcgttc ctgctgctgc tcatcgtcgt tggtcaccgc 60
gcc 63

<210> 30
<211> 63
<212> DNA
<213> Artificial Sequence

<220>
<223> D-Hordian signal peptide sequence

<400> 30
atggctaagc ggctggtcct ctttgtggcg gtaatcgtcg ccctcgtggc tctcaccacc 60
gcc 63

<210> 31
<211> 1314
<212> DNA
<213> Artificial Sequence

<220>
<223> O2 transcription factor sequence

<400> 31
atggagcacg tcatctcaat ggaggagatc ctcgggccct tctgggagct gctaccaccg 60
ccagcgccag agccagagcg agagcagcct ccggtaaccg gcatcgtcgt cggcagtgtc 120
atagacgttg ctgctgctgg tcatggtgac ggggacatga tggatcagca gcacgccaca 180
gagtggacct ttgagaggtt actagaagag gaggctctga cgacaagcac accgccgccg 240
gtggtggtgg tgccgaactc ttgttgctca ggcgccctaa atgctgaccg gccgccggtg 300
atggaagagg cggtaactat ggcgcctgcg gcggtgagta gtgccgtagt aggtgacccc 360
atggagtaca atgccatact gaggaggaag ctggaggagg acctcgaggc cttcaaaatg 420
tggagggcgg cctccagtgt tgtgacctca gatcaacgtt ctcaaggctc aaacaatcac 480
actggaggta gcagcatcag gaataatcca gtgcagaaca agctgatgaa cggcgaagat 540
ccaatcaaca ataaccacgc tcaaactgca ggccttggcg tgaggcttgc tactagctct 600
tcctcgagag atccttcacc atcagacgaa gacatggacg gagaagtaga gattctgggg 660
ttcaagatgc ctaccgagga aagagtgagg aaaagaaagg aatccaatag agaatcagcc 720
agacgctcga gatacaggaa agccgctcac ctgaaagaac tggaagacca ggtagcacag 780
ctaaaagccg agaattcttg cctgctgagg cgcattgccg ctctgaacca gaagtacaac 840
gacgctaacg tcgacaacag ggtgctgaga gcggacatgg agaccctaag agctaaggtg 900
aagatgggag aggactctct gaagcgggtg atagagatga gctcatcagt gccgtcgtcc 960
atgcccatct cggcgccgac ccccagctcc gacgctccag tgccgccgcc gcctatccga 1020
gacagcatcg tcggctactt ctccgccaca gccgcagacg acgatgcttc ggtcggcaac 1080
ggtttcttgc gactgcaagc tcatcaagag cctgcatcca tggtcgtcgg tggaactctg 1140
agcgccacag agatgaaccg agtagcagca gccacgcatt gcgcgggggc catggagcac 1200
atccagacgg cgatgggatc catgccgccg acctccgcct ccggatctac accgccgccg 1260
caggattatg agctgctggg tccaaatggg gccatacaca tggacatgta ttag 1314

<210> 32
<211> 987
<212> DNA
<213> Artificial Sequence

<220>
<223> PBF transcription factor sequence

<400> 32
atggacatga tctccggcag cactgcagca acatcaacac cccacaacaa ccaacaggcg 60
gtgatgttgt catcccccat tataaaggag gaagctaggg acccaaagca gacacgagcc 120

atgccccaaa taggtggcag tggggagcgt aagccgaggc cgcaactacc tgaggcgctc 180
aagtgcccac gctgcgactc caacaacacc aagttttgct actacaacaa ttatagcatg 240
tcacaaccac gctacttttg caaggcttgc cgccgctatt ggacacatgg tggtaccctc 300
cgcaatgtcc ccattggtgg tgggtgtcgc aagaacaaac atgcctctag atttgtcttg 360
ggctctcaca cctcatcgtc ctcatctgct acctatgcac cattatcccc tagcaccaac 420
gctagctcta gcaatatgag catcaacaaa catatgatga tggtgcctaa catgacgatg 480
cctaccccaa cgacaatggg cttattccct aatgtgctcc caacacttat gccgacaggt 540
ggaggcgggg gctttgactt cactatggac aaccaacata gatcattgtc cttcacacca 600
atgtctctac ctagccaggg gccagtgcct atgctggctg caggagggag tgaggcaaca 660
ccgtctttcc tagagatgct gagaggaggg atttttcatg gtagtagtag ctataacaca 720
agtctcacga tgagtggtgg caacaatgga atggacaagc cattttcgct gccatcatat 780
ggtgcaatgt gcacaaatgg gttgagtggc tcaaccacta atgatgccag acaactggtg 840
gggcctcagc aggataacaa ggccatcatg aagagcagta ataacaacaa tggtgtatca 900
ttgttgaacc tctactggaa caagcacaac aacaacaaca acaacaacaa caacaacaac 960
aacaacaaca acaacaaggg acaataa 987

<210> 33
<211> 3902
<212> DNA
<213> Artificial Sequence

<220>
<223> Reb transcription factor sequence

<400> 33
atggagcggg tgttctccgt ggaggagatc tccgacccat tctgggtccc gcctccgccg 60
ccgcagtcgg cggcggcggc ccagcagcag ggcggcggcg gcgtggcttc gggaggtggt 120
ggtggtgtag cgggggcgcg cggcggcggg aacgcgatga accggtgccc gtcggagtgg 180
tacttccaga agtttctgga ggaggcggtg ctcgatagcc ccgtcccgaa ccctagcccg 240
agggccgaag cgggagggat caggggcgca ggaggggtgg tgccggtcga tgttaagcag 300
ccgcagctct cggcggcggc gacgacgagc gcggtggtgg accccgtgga gtacaacgcg 360
atgctgaagc agaagctgga gaaggacctc gccgcggtcg ccatgtggag ggtacagcca 420
ttctcccccc ctctagtact cgagagctta ctgagatcgg caatgctagc tactgtttgc 480
atcgaatgtt tataggtatt tagatcgggc atttctatag accaatggcg tccatggtct 540
tgcaatgcgc tctgttgagt gtcggtggtt ggttcgactc atagtatgta gggttgtgcg 600
tatgtacaaa cggaagcttc atagacctcg gtattgagat tgcgatatcg atgcaacctg 660
cgaattggcg atgtaatcag tcatattctt actaaactgc gagacagtgg tttgtttgca 720
attgcaatat ttttgtatgg ggctgcttaa actgtcattg ccttttttaga ttggcaatat 780
gtgactttat gcaagtattt gattgggcgg atccaggaac aaaaagttgg ggggattcaa 840
cataccgagt acactggcat aaacacatca tctcagtatt aaactatgct aaaatgctat 900
taagagacct ttagcacctc ttatcttatc aaccatggtg aaaaaattga agggggggact 960
cagggggta tccatgggtc cgatgggtgc agggggggact gagtcccccc tgcacccacg 1020
ttgaatccgc cctggcatgc gtataagctg tcacagccat ttctaggtgc ttgtgcttag 1080
ttgggtgatg tcagcttaat ttgtcttttc tatgtcgtca tcgattttct aagaaacgaa 1140
aaatagccta tttatgtgct ccagaatttg atgatccctg gcccttcatt tgctgaaatt 1200
agcctatttg ttggttgccc ttcagttttt tcccagctta tgttgttgca atgtgtggct 1260
atgcctcgtt ttgtgcccta taatttatta tttgcaattc atttttgtac atgacttaaa 1320
atgacactag agcaacatgc actgattggt tatcctataa tcatttatgt agttctgttc 1380
attttatcat gctagctcat gtcattttca tcttcaggcc tctggcacag ttccacctga 1440
gcgtcctgga gctggttcat ccttgctgaa tgcagatgtt tcacacatag gcgctcctaa 1500
ttccatcgga ggtacttatc ttatctggtt acattttcag attgttatga aactacccaa 1560
atatcctgca caattgcatg ggattaaatt ttagtttctt tgaaatagaa gtagagttgt 1620
attgctgtca cgtcatcaaa tagttctgaa gctatgaata aataagttcc gcatttgtta 1680
gtgattcttt gaacattaga attgttatgc ttaagtagat agggttatgt ttgtttggag 1740
ttcccttaaa tcatttcatt gctgactgcc agctggcagg agcatttgtt gttgccttga 1800
ccatgaatga agaccttcct gttctgagtg ctcacaagaa aacatatttt gattaatgca 1860
ccttgaatcc ttaggatctt gcaaagatgg gcacttagct ttagaattga gtagtactta 1920
aatagctgtt gttatcatga tttgtcctgt agtgaaatgt cgacaaaaca ggaatgctac 1980
ttttgacttc tgatatttca tgcctggctt tacttatgct ctgtttggaa catgggcaca 2040
tatcaggcaa tgctactcca gttcaaaaca tgctaagtgg cccaagtggg ggatcgggct 2100
cacagttggt acagaatgtt gatgtccttg taaagcagcc caccagctct tcatcaaggg 2160
agcagtcaga tgatgatgac atgaagggag aagctgagac cactggaact gcaagacctg 2220
ctgatcaaag attacaacga aggtgatcat tcattgcttc cttgtaatat agattctgta 2280
cataattaac ctacctcgtc atgcatgcat gtgtcctatt ttcaccttag ccctttcagt 2340
tggatttcca ctttcatccg gtagcctttc agtttcctat tgcatcgcat atatgatctt 2400

ttacctacca tattagttct ctgtgtgcca tactcagtgc ttagtgtctc gagcaagaga 2460
ggaatttgta tggctattac acgtagcact ttgctctcta cttgtttatt gacataagca 2520
atttgggatg aattaaatct gagttcacat catattcctt atgtcacaag tttctgaaac 2580
cgattgtatc tagtatctgg ttgatgcacc cccatcttgg atttgcaaat caaagttata 2640
ctccctagag agctttacct ttcataaagc aattacccca ataaaccacg gatttgatag 2700
ctattgacta tgattaccag aattcatttg gcagctattt tctcaattta agtttggtat 2760
tagtctcagt tggctgtaaa ataatgtcac ggtagggtac atgtatgtgc agcatacaag 2820
gtatgggtga gttatgatat ggacagtgtg tacaccccac atttgctcac taaaatcaaa 2880
atattcaaac gtcacgtgat gatatggtgg attgcattat accttgtatt gtttattatg 2940
ttacttgtgc tagacaataa tataggctgt tcttttgggt gattttgtat gaagatgttg 3000
agcaagcact tctcgatata atgctagttt tgttgacctg ttccaggaag caatccaatc 3060
gggagtcagc caggcgctca agaagcagaa aggcagctca cttgaatgag ctggaggcac 3120
aggtgtgata gttcacatag ttattttcga taagacataa aatcctaaat tactggctac 3180
tgacttcagt tatggattta cttgttacag gtatcgcaat taagagtcga gaactcctcg 3240
ctgttaaggc gtcttgctga tgttaaccag aagtacaatg atgctgctgt tgacaataga 3300
gtgctaaaag cagatgttga gaccttgaga gcaaaggtat gctatatatg ccttttgcaa 3360
tatgcatccc atggattgct actttggctt gtttcaaact ttcaacgtga cttgtgtacc 3420
ctgttattag aagaataatc ccgcctacca ttatactcta taaatcacca tttggccagt 3480
ccaaacatga ttattaaatc aggtcaatct gaacattgaa atgtatcaaa aattcgcagg 3540
tgaagatggc agaggactcg gtgaagcggg tgacaggcat gaacgcgttg tttcccgccg 3600
cttctgatat gtcatccctc agcatgccat tcaacagctc cccatctgaa gcaacgtcag 3660
acgctgctgt tcccatccaa gatgacccga acaattactt cgctactaac aacgacatcg 3720
gaggtaacaa caactacatg cccgacatac cttcttcggc tcaggaggac gaggacttcg 3780
tcaatggcgc tctggctgcc ggcaagattg gccggccagc ctcgctgcag cgggtggcga 3840
gcctggagca tctccagaag aggatgtgcg gtgggccggc ttcgtctggg tcgacgtcct 3900
ga 3902

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2002 (22.08.2002)

PCT

(10) International Publication Number
WO 02/064750 A3

(51) International Patent Classification[7]: A01H 5/00, C12N 15/82, 15/87

(21) International Application Number: PCT/US02/04909

(22) International Filing Date: 14 February 2002 (14.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/269,199	14 February 2001 (14.02.2001)	US
60/269,188	14 February 2001 (14.02.2001)	US
09/847,232	2 May 2001 (02.05.2001)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/847,232 (CIP)
Filed on 2 May 2001 (02.05.2001)

(71) Applicant *(for all designated States except US)*: **VENTRIA BIOSCIENCE** [US/US]; 4110 North Freeway Boulevard, Sacramento, CA 95834-1219 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HUANG, Ning** [CN/US]; 1226 Caricia Drive, Davis, CA 95616 (US). **YANG, Daichang** [CN/US]; 1318 F Street, #3, Davis, CA 95616 (US).

(74) Agents: **ARNETT, Stephen, E.** et al.; Perkins Coie LLP, P.O. Box 2168, Menlo Park, CA 94026 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
21 November 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: EXPRESSION SYSTEM FOR SEED PROTEINS

(57) Abstract: The invention is directed to methods and compositions for high level expression of heterologous polypeptides in the grains of transgenic plants and compositions resulting therefrom that are suitable for oral delivery or other uses.

WO 02/064750 A3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/04909

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : A01H 5/00; C12N 15/82, 15/87
US CL : 435/320.1; 800/278,287,298,288
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/320.1; 800/278,287,298,288

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EAST(USPAT,EPO,JPO,DERWENT); STN(AGRICOLA,BIOSIS,CAPLUS,EMBASE)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WU et al. The GCN4 motif in a rice glutelin gene is essential for endosperm-specific gene expression and is activated by Opaque-2 in transgenic rice plants. The Plant Journal, 1998, Vol. 14, No. 6, pages 673-683, see especially pages 677-679.	1,3-6,8-23,26,28-36,47,48,53,56-69,118 and 119
Y	US 5,543,576 A (VAN OOIJEN ET AL) 6 August 1996 (06.08.1996), see columns 1-3.	1,3-6,8-23,26,28-36,47,48,53,56-69,118 and 119
Y	HORVATH et al, The production of recombinant proteins in transgenic barley grains. PNAS, 15 February 2000, Vol. 97, No. 4, pages 1914-1919, see entire document, especially Figure 2 on page 1915.	1,3-6,8-23,26,28-36,47,48,53,56-69,118 and 119
Y	ZHENG et al, A distal promoter region of the rice seed storage protein glutelin gene enhanced quantitative gene expression. Plant Science, 1997, Vol. 128, pages 59-65, see entire document.	1,3-6,8-23,26,28-36,47,48,53,56-69,118 and 119

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 September 2002 (11.09.2002)	27 SEP 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer. Felicia D. Roberts for David H Kruse Telephone No. 703-308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/04909

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claim Nos.: 25,37-45,49-52 and 77-79
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: 1,3-6,8-23,26,28-36,47,48,53,56-69,118 and 119

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.
☒ No protest accompanied the payment of additional search fees.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions, which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1, 3-6, 8-23, 26, 28-36, 47, 48 and 118-119, drawn to a method of producing a heterologous polypeptide in a grain of a plant comprising a chimeric gene comprising a rice glutelin Gt1 promoter, said chimeric gene, an expression vector comprising said chimeric gene and a transformed host and its parts comprising said chimeric gene.

Group II, claim(s) 2-22, 24, 27-36, 46, 48, 118 and 119, drawn to a method of producing a heterologous polypeptide in a grain of a plant comprising a chimeric gene comprising a rice glutelin Glb promoter, said chimeric gene, an expression vector comprising said chimeric gene and a transformed host and its parts comprising said chimeric gene.

Group III, claim(s) 54 and 55, drawn to a codon optimised nucleic acid molecule.

Group IV, claim(s) 56-69, drawn to a method of producing a heterologous polypeptide in a grain of a plant comprising a chimeric gene comprising a promoter of a storage protein gene.

Group V, claim(s) 70-72, drawn to an expression vector comprising an aleurone-specific promoter.

Group VI, claim(s) 73-76 drawn to a transformed host comprising a chimeric gene comprising an aleurone-specific promoter,

Group VII, claim(s) 80-89, drawn to a method of processing a transgenic seed that comprises a heterologous polypeptide.

Group VIII, claim(s) 90-92 and 95-97, drawn to a method of processing transgenic grains containing a heterologous polypeptide to produce a malt syrup.

Group IX, claim(s) 93-97, drawn to a method of processing transgenic grains containing a heterologous polypeptide to produce a malt extract.

Group X, claim(s) 98-110, drawn to a transgenic malt syrup or malt extract, various products comprising said syrup or extract and various products comprising a protein-containing product.

Group XI, claim(s) 111-117, drawn to a method for producing a selected protein other than a monocot-plant seed-storage protein.

The inventions listed as Groups I-XI do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The broadest general concept of the instant invention can be found in Group IV, claims 56-69, as a method of producing a heterologous polypeptide in a grain of a plant comprising a chimeric gene comprising a promoter of a storage protein, a nucleic acid leader encoding a signal peptide of a storage protein and a heterologous nucleic acid that encodes the heterologous polypeptide. This general concept was known in the art prior to applicant's invention and thus does not make a contribution over the prior art as required under PCT Rule 13.2. See, for example, Vanderkerckhove *et al* (U.S. Patent 5,487,991, published 30 January 1996) especially the claims. Hence, the instant claims as a whole do not relate to one invention or group of inventions so linked as to form a single general inventive concept as required under PCT Rule 13.1, nor do all of the claimed inventions contribute over the prior art as required under PCT Rule 13.2.

In addition, the chimeric gene and the expression vector of Group I cannot be used in the method of Group II, and the optimized nucleic acid molecule of Group III, a third product, is not required by the methods of Group I or Group II. The invention of Group IV, a third method requires promotes not requied by any other group. Group V, drawn to a fourth product, requires a promoter not required by any other group. Group VI, a fifth product, requires microbial hosts not required by any other group. Groups VII-IX,

drawn to fourth through sixth methods, require particular processing steps and reagents not required by each other or any other group. Group X, a sixth product, requires chemical components distinct from any other group. Group XI, a seventh method, requires monocot transformation and protein extraction methods not required by the other groups.